

RESUM

Aquest Treball de Fi de Grau abasta la creació d'una aplicació web per a la consulta d'informació estadística dels resultats acadèmics de les assignatures de primer any del Grau en Enginyeria en Tecnologies Industrials de l'Escola Tècnica Superior d'Enginyeria Industrial de Barcelona, de la Universitat Politècnica de Catalunya.

L'aplicació permet de consultar la previsió de les notes del segon quadrimestre d'un estudiant basant-se amb les qualificacions del anterior quadrimestre i amb la nota d'accés a la universitat. A més, l'aplicació també retorna, com a mesura de marge d'error, l'interval de confiança del 95% d'aquesta nota estimada.

S'opta per emprar la regressió lineal múltiple com a mètode de predicció estadística, i utilitzar el software estadístic R per al tractament de les dades. Es busquen els corresponents algoritmes necessaris per a filtrar les dades necessàries d'aquelles que no són d'interès. Seguidament es dura a terme la metodologia de l'eliminació progressiva per a trobar els models òptims de regressió lineal múltiple.

Tanmateix, es decideix de fer servir els llenguatges informàtics HTML, CSS i JavaScript per a la modelització de la interfície d'usuari de l'aplicació i fer que aquesta sigui operativa.

Finalment, s'exposen la planificació temporal, els costos associats al projecte i s'avalua el seu impacte social i ambiental.

SUMARI

RESUM	1
LLISTAT DE FIGURES.....	5
LLISTAT DE TAULES.....	6
GLOSSARI	7
Paràmetres	7
Sigles.....	7
1. PREFACI	8
1.1. Motivació.....	8
1.2. Requeriments previs	8
2. INTRODUCCIÓ.....	9
2.1. Objectius del projecte	9
2.2. Estructura del projecte.....	9
2.3. Abast del projecte.....	10
3. PUNT DE PARTIDA.....	11
3.1. Els fitxers inicials amb les dades acadèmiques.....	11
3.2. El disseny de la interfície d'usuari de l'aplicació.....	13
3.3. La programació estadística i la programació web.	14
3.3.1. La programació estadística (R)	14
3.3.2. La programació web (HTML, CSS, JavaScript)	15
4. LA REGRESSIÓ LINEAL MÚLTIPLE.....	16
4.1. Introducció matemàtica de la regressió lineal múltiple	16
4.2. La comprovació del model	18
4.3. El mètode de l'eliminació progressiva (<i>Backward Selection</i>)	20
4.3.1. El mètode tradicional.....	20
4.3.2. El mètode modern	21
5. FILTRAT DE LES DADES	22
5.1. Dades d'interès.....	22
5.2. L'algoritme del filtrat de les dades	22
5.3. L'anàlisi en R de la regressió lineal múltiple	25
5.3.1. Anàlisi de correlacions entre les variables explicatives.....	25
5.3.2. Realització dels models inicials de regressió lineal múltiple.....	27
5.3.3. Comprovació dels supòsits dels models.....	27

5.3.4. Realització dels models finals de la regressió lineal múltiple (<i>Backward Selection</i>)	32
5.4. El models definitius.....	39
6. L'APLICACIÓ WEB	41
6.1. La interfície d'usuari	41
6.2. El JavaScript.....	41
7. PROPOSTA DE CONTINUACIÓ DEL TREBALL.....	46
8. PLANIFICACIÓ TEMPORAL I COSTOS	47
8.1. Planificació temporal	47
8.2. Costos.....	48
9. IMPACTE SOCIAL I AMBIENTAL.....	50
9.1. Impacte social.....	50
9.2. Impacte ambiental	50
CONCLUSIONS.....	51
AGRAÏMENTS.....	52
BIBLIOGRAFIA	53
Referències bibliogràfiques	53
Bibliografia complementària.....	53

LLISTAT DE FIGURES

Figura 2.1. Diagrama de blocs de l'estructura del projecte.	10
Figura 3.1. Dades del fitxer Dades_totes_qualif1_erfins2016.csv.	11
Figura 3.2. Dades del fitxer Dades_pers_IND_nomes_preins2016.csv.	12
Figura 3.3. Disseny esquemàtic de la interfície d'usuari principal de l'aplicació.	13
Figura 3.4. Disseny esquemàtic de la interfície d'usuari principal de l'aplicació després de realitzar una consulta.	14
Figura 4.1. Exemple de gràfic qq en el qual es verifica la normalitat.	18
Figura 4.2. Exemple de verificació d'homogeneïtat constant dels residus.	19
Figura 4.3. Exemple de verificació de l'efecte lineal d'una variable explicativa.	19
Figura 5.1. Crida dels fitxers inicials de dades i creació de F1 i F2.	23
Figura 5.2. Creació de la matriu FF.	24
Figura 5.3. Creació de la matriu Dades.	25
Figura 5.4. ScatterplotMatrix i boxplots de les variables explicatives.	26
Figura 5.5. Gràfics qq dels residus dels models inicials.	28
Figura 5.6. Gràfics dels residus estandaritzats contra els valors estimats pel model.	29
Figura 5.7. Gràfics de les variables explicatives pel model de Geometria.	30
Figura 5.8. Gràfics de les variables explicatives pel model de Càlcul II.	30
Figura 5.9. Gràfics de les variables explicatives pel model de Termodinàmica.	31
Figura 5.10. Gràfics de les variables explicatives pel model de Química II.	31
Figura 5.11. Gràfics de les variables explicatives pel model d'Expressió Gràfica.	32
Figura 6.1. Interfície d'usuari de l'aplicació.	41
Figura 6.2. Codi de la Funcio1.	42
Figura 6.3. La interfície d'usuari en el cas d'una nota fora de rang.	43
Figura 6.4. La interfície d'usuari en el cas de totes les notes dins de rang.	43
Figura 6.5. Primer part del codi de la Funcio2.	44
Figura 6.6. Segona part del codi de la Funcio2.	45
Figura 8.1. Diagrama de Gantt del projecte.	47

LLISTAT DE TAULES

Taula 4.1. Taula genèrica de dades.	16
Taula 5.1. AIC del model inicial de Geometria i del model sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$	33
Taula 5.2. AIC del model actual de Geometria i del model sense cada una de les variables, i el p-valor dels coeficients del model actual $C'i = 0$	33
Taula 5.3. AIC del model inicial de Càlcul II i l'AIC d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$	34
Taula 5.4. AIC del model actual de Càlcul II i del model sense cada una de les variables, i el p-valor dels coeficients del model actual $C'i = 0$	34
Taula 5.5. AIC del model actual de Càlcul II i del model sense cada una de les variables, i el p-valor dels coeficients del model actual $C'i = 0$	35
Taula 5.6. AIC del model inicial de Termodinàmica Fonamental i l'AIC d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$	35
Taula 5.7. AIC del model actual de Termodinàmica Fonamental i d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$	36
Taula 5.8. AIC del model inicial de Química II i l'AIC d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$	36
Taula 5.9. AIC del model actual de Química II i d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$	37
Taula 5.10. AIC del model actual de Química II i d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$	37
Taula 5.11. AIC del model inicial d'Expressió Gràfica i l'AIC d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$	38
Taula 5.12. AIC del model actual d'Expressió Gràfica i d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$	38
Taula 5.13. Coeficients del model final de Geometria.	39
Taula 5.14. Coeficients del model final de Càlcul II.	39
Taula 5.15. Coeficients del model final de Termodinàmica Fonamental.	39
Taula 5.16. Coeficients del model final de Química II.	39
Taula 5.17. Coeficients del model final d'Expressió Gràfica.	40
Taula 8.1. Taula resum de les despeses.	49

GLOSSARI

Paràmetres

β	Vector columna que té com a components els coeficients $\beta_0, \beta_1, \dots, \beta_p$ d'un model de regressió lineal múltiple.
ε	Vector columna els components ε_i de la qual segueixen una seqüència aleatòria amb mitjana zero i variància σ^2 .
σ^2	Variància de $\varepsilon_i \sim N(0, \sigma^2)$.
$t_{n,m}$	És el valor del quantil n de la distribució t de Student amb m graus de llibertat.
α	Significança estadística
K	Nombre de paràmetres d'un model.
L	Màxima versemblança d'un model.
r	Coeficient de correlació de Pearson.

Sigles

ETSEIB	Escola Tècnica Superior d'Enginyeria Industrial de Barcelona
CSV	<i>comma-separated values</i>
AIC	<i>Akaike Information Criterion</i>
ECTS	<i>European Credit Transfer and Accumulation System</i>
HTML	<i>Hypertext Markup Language</i>
CSS	<i>Cascading Style Sheets</i>
IC	Interval de Confiança
IP	Interval de Predicció

1. PREFACI

1.1. Motivació

El gran interès desenvolupat al llarg de les assignatures de Fonaments d'Informàtica, d'Informàtica i d'Estadística va ser la principal motivació per a dur a terme aquest multidisciplinari projecte en l'àmbit de la programació i d'estadística.

Així mateix, poder posar en pràctica els coneixements adquirits en aquests àmbits, així com les competències i capacitats desenvolupades al llarg del grau, per a la realització d'un projecte, resulta atractiu i suggestiu, al mateix temps que enriquidor a nivell personal.

1.2. Requeriments previs

Els requeriments previs necessaris per a la realització d'aquest projecte són els coneixements obtinguts a les assignatures de Fonaments d'Informàtica i d'Estadística. A més, és necessari de desenvolupar, previ al projecte, una base sòlida de coneixement en llenguatge de programació web.

2. INTRODUCCIÓ

2.1. Objectius del projecte

L'objectiu principal del projecte és la creació d'una aplicació web per a mòbil pel tractament estadístic de resultats acadèmics. L'aplicació estarà enfocada a l'ús d'un estudiant que vol conèixer quina assignatura de les que ha d'afrontar li pot suposar més dificultat. Aquest objectiu s'aconsegueix mitjançant una predicció estadística de la nota esperada de l'usuari a l'assignatura que es sol·liciti informació.

2.2. Estructura del projecte

A fi d'assolir els objectius, s'enfoca el treball des de dos dominis interconnectats entre si: la informàtica i l'estadística. Es partirà de dos fitxers en format .csv amb totes les dades acadèmiques necessàries.

En primer lloc, es procedirà a la tria tant del software estadístic (R), com dels llenguatges programació web (HTML, CSS, JavaScript) amb els quals s'enfocarà el treball. A continuació, s'elaborarà un algoritme en llenguatge R que permeti filtrar les dades per a netejar aquelles que no resultin necessàries (bé perquè no són d'interès per l'estudi, bé perquè les dades estan repetides, etc.).

També s'usarà R per a obtenir els models de regressió lineal múltiple respectius a cada assignatura. Caldrà comprovar les suposicions de les qual parteixen els models. Així mateix s'emprarà el mètode de l'eliminació progressiva a fi d'aconseguir que els models de regressió múltiple siguin òptims.

Finalment, es procedirà al disseny i creació de la interfície d'usuari de l'APP i a la interconnexió d'aquesta amb l'anàlisi estadístic anterior.

A la següent figura es mostra el diagrama de blocs amb l'estructura del projecte.

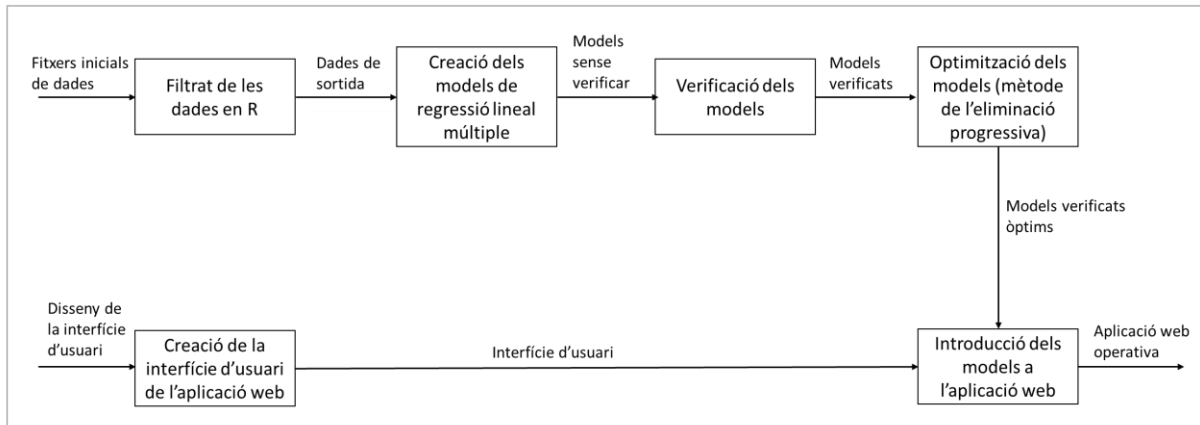


Figura 2.1. Diagrama de blocs de l'estructura del projecte.

2.3. Abast del projecte

Aquest treball pretén recollir l'escènica del que podria ser un treball més ampli i extens. Es realitzarà un enfoc estadístic destinat especialment a l'ús d'un usuari amb el perfil d'un estudiant del Grau en Enginyeria en Tecnologies Industrials de l'Escola Tècnica Superior d'Enginyeria Industrial de Barcelona (ETSEIB) de primer any. Així doncs, no es tractaran dades referents a estudiants d'altres cursos, d'altres graus o d'altres facultats. A més, es restringirà a aquells estudiants que han entrat mitjançant les Proves d'Accés a la Universitat per a estudiants provinents de Batxillerat.

3. PUNT DE PARTIDA

3.1. Els fitxers inicials amb les dades acadèmiques

Es disposa inicialment de dos fitxers en format csv amb les dades acadèmiques necessàries per a l'estudi proposat. El format csv (de l'angles *comma-separated values*) és una tipologia de documents emprada per a la representació de les dades en forma de matriu o taula.

Tanmateix, el primer fitxer (Dadestotesqualif1erfins2016.csv) és compost per 40986 files i 11 columnes. Cada fila conté la informació d'una convocatòria d'una assignatura d'un alumne de l'ETSEIB. L'arxiu recull totes les convocatòries, assignatures i estudiants fins el 2015 inclòs.

La primera columna indica el codi del programa, on 752 és el Grau en Enginyeria en Tecnologies Industrials. La segona columna conté el codi de l'expedient de l'alumne, la tercera el codi de l'assignatura i la cinquena el nombre de crèdits ECTS corresponents a l'assignatura. La sisena i la setena columna indiquen respectivament l'any i el quadrimestre de la convocatòria. La setena columna informa de si l'estudiant ha superat o no l'assignatura, i la vuitena, la novena i la desena columna mostren diferents criteris d'avaluació. Finalment, la onzena columna indica del grup classe de l'estudiant.

	CODI_PROGRAMA	CODI_EXPEDIENT	CODI_UPC_UD	CREDITS	CURS	QUAD	SUPERA	NOTA_PROF	NOTA_NUM_AVAL	NOTA_NUM_DEF	GRUP_CLASSE
1	752	229928	240015	6	2010	1	S				CONV
2	752	229036	240013	6	2010	1	S				CONV
3	753	230399	240011	6	2010	1	S				CONV
4	752	232924	240014	6	2010	1	S				CONV
5	753	232121	240214	6	2010	1	S				CONV
6	752	232295	240014	6	2010	1	S	6,5	6,5	6,5	CONV
7	752	228100	240012	6	2010	2	N	0	0	0	21
8	752	228382	240012	6	2010	2	N	4,4	4,4	4,4	11
9	753	231333	240012	6	2010	2	N	0	0	0	21
10	754	232820	240012	6	2010	2	N	3,9	3,9	3,9	21
11	752	227236	240012	6	2010	2	N	1,5	1,5	1,5	11
12	752	229262	240012	6	2010	2	S	5	5	5	11
13	752	227979	240012	6	2010	2	N	4,6	4,6	4,6	11
14	752	227320	240012	6	2010	2	N	4	4	4	11
15	752	226888	240012	6	2010	2	N	4,5	4,5	4,5	12
16	752	228704	240012	6	2010	2	S	5,4	5,4	5,4	12
17	752	231325	240012	6	2010	2	S	5,2	5,2	5,2	12
40972	752	289378	240021	6	2015	1	N	4,6	4,6	4,6	10
40973	752	289398	240021	6	2015	1	S	4	4	5	10
40974	752	289405	240021	6	2015	1	S	5,4	5,4	5,4	10
40975	752	304279	240015	6	2015	1	N	4,1	4,1	4,1	91
40976	752	304393	240015	6	2015	1	S	6,5	6,5	6,5	92
40977	752	308061	240015	6	2015	1	N	3,1	3,1	3,1	101
40978	752	303994	240015	6	2015	1	S	9,1	9,1	9,1	23
40979	752	289389	240015	6	2015	1	S	5,5	5,5	5,5	23
40980	752	304333	240015	6	2015	1	S	9,2	9,2	9,2	13
40981	752	309128	240015	6	2015	1	N	1,6	1,6	1,6	21
40982	752	304354	240015	6	2015	1	S	9	9	9	13
40983	752	303959	240015	6	2015	1	N	1,9	1,9	1,9	13
40984	752	304246	240015	6	2015	1	N	2,6	2,6	2,6	13
40985	752	304103	240015	6	2015	1	S	9,7	9,7	9,7	23
40986	752	303960	240015	6	2015	1	S	10	10	10	23

Figura 3.1. Dades del fitxer Dadestotesqualif1erfins2016.csv.

D'altra banda, el segon fitxer (DadespersINDnombrespreins2016.csv) conté dades personals i acadèmiques preuniversitàries dels estudiants del Grau en Enginyeria en Tecnologies Industrials de totes les convocatòries fins 2015 (inclòs). Està compost per a 2808 files, on cada fila inclou la informació respectiva a un estudiant, i 9 columnes.

La primera columna indica el codi del programa i per tant pren el mateix valor 752 per a tots els estudiants. La segona, tercera i quarta columna contenen respectivament el codi de l'estudiant, el sexe i el codi postal familiar. La cinquena columna mostra l'any d'accés, i la sisena i setena columna indiquen respectivament el tipus i la via d'accés (aquest valors són constants per a totes les files). Finalment, la vuitena conté la nota d'admissió a la universitat, i la darrera columna conte el codi postal del centre escolar.

	CODI_PROGRAMA	CODI_EXPEDIENT	SEXE	CP_FAMILIAR	ANY_ACCES	TIPUS_ACCES	VIA_ACCES	NOTA_ACCES	CP_CENTRE_SEC
1	752	277551	H	08328	2013	1	0	12.042	8328
2	752	244848	H	25790	2011	1	0	11.390	25280
3	752	245234	H	08017	2011	1	0	10.240	8017
4	752	227703	D	08037	2010	1	0	10.586	8010
5	752	245776	H	08019	2011	1	0	9.774	8009
6	752	228122	H	08001	2010	1	0	9.772	8034
7	752	245810	H	08506	2011	1	0	11.400	8500
8	752	244262	H	43500	2011	1	0	10.264	43500
9	752	244287	H	43500	2011	1	0	11.204	43500
10	752	244289	H	43850	2011	1	0	12.008	43850
11	752	244235	H	43205	2011	1	0	12.330	43205
12	752	294152	H	08172	2014	1	0	12.014	8173
13	752	227686	H	08902	2010	1	0	10.076	8014
14	752	231199	H	25002	2010	1	0	11.330	25002
15	752	245591	H	08840	2011	1	0	10.294	8007
16	752	245620	D	08014	2011	1	0	10.374	8014
17	752	245637	H	08021	2011	1	0	10.004	8034
2795	752	304166	H	08021	2015	1	0	10.076	8017
2796	752	304020	H	08006	2015	1	0	12.742	8022
2797	752	304022	D	17003	2015	1	0	12.516	17003
2798	752	308055	H	12002	2015	1	0	10.360	NA
2799	752	308065	H	08005	2015	1	0	12.230	8010
2800	752	304396	H	08029	2015	1	0	11.210	8029
2801	752	304398	D	08036	2015	1	0	12.444	8017
2802	752	304399	H	08025	2015	1	0	9.870	8025
2803	752	304402	H	35420	2015	1	0	11.245	NA
2804	752	304035	H	08024	2015	1	0	10.972	8034
2805	752	304037	H	08022	2015	1	0	10.828	8034
2806	752	304125	H	08021	2015	1	0	11.278	8022
2807	752	304049	H	07210	2015	1	0	11.444	NA
2808	752	304069	H	08840	2015	1	0	10.918	8840

Figura 3.2. Dades del fitxer DadespersINDnombrespreins2016.csv.

3.2. El disseny de la interfície d'usuari de l'aplicació

L'aplicació ha de permetre, donades les notes del primer quadrimestre i la nota d'accés a la universitat, mostrar una predicció estadística de la nota esperada de l'usuari en l'assignatura que aquest sol·liciti.

Així doncs, l'usuari es trobarà en una primera interfície en la que se li demanarà la nota d'accés a la universitat i les qualificacions de les assignatures del primer quadrimestre: Àlgebra Lineal, Càlcul I, Mecànica Fonamental, Química I i Fonaments d'Informàtica. També se li demanarà de quina assignatura del segon quadrimestre vol conèixer la informació. Les assignatures del segon quadrimestre són: Geometria, Càlcul II, Termodinàmica Fonamental, Química II i Expressió Gràfica.

Es mostra a continuació una representació esquemàtica de la interfície d'usuari de l'aplicació:

Diagrama esquemàtic de la interfície d'usuari principal de l'aplicació:

- Títol:** Consulta d'estadístiques de resultats acadèmics
- Text informatiu:** Instruccions i indicacions del formulari
- Formulari:**
 - Àlgebra Lineal
 - Càlcul I
 - Mecànica Fonamental
 - Química I
 - Fonaments d'Informàtica
 - Nota d'Accés a la Universitat
 - Assignatura a preveure (menú desplegable i botó Continuar)

Figura 3.3. Disseny esquemàtic de la interfície d'usuari principal de l'aplicació.

Posteriorment, l'usuari observarà la seva nota prevista. Tanmateix, s'oferirà a l'usuari la possibilitat de canviar qualsevol paràmetre dels omplerts prèviament i actualitzar el contingut consegüentment.

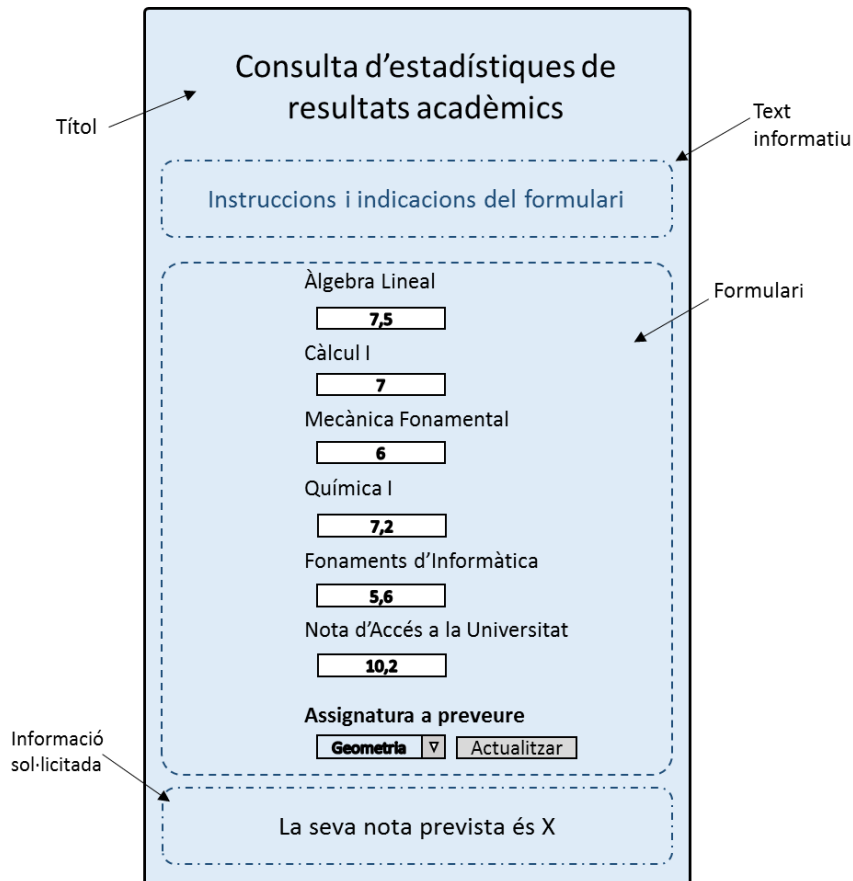


Figura 3.4. Disseny esquemàtic de la interfície d'usuari principal de l'aplicació després de realitzar una consulta.

3.3. La programació estadística i la programació web.

3.3.1. La programació estadística (R)

El codi de programació estadístic emprat és R: un *software* gratuït enfocat a l'estadística computacional i de gràfics. R és disponible per sistemes operatius Unix, Linux, Windows i MacOS. És el codi de programació líder per a l'estadística i l'anàlisi de dades, amb més de 4000 desenvolupadors a nivell mundial. A més, és capdavanter en la realització de gràfics estadístics. (1)

3.3.2. La programació web (HTML, CSS, JavaScript)

Un servidor web és un *software* que rep les sol·licitacions d'una pàgina web i respon a aquestes. Un navegador web mostra els documents dels servidors web. Així mateix, cada pàgina web és programada mitjançant un llenguatge de programació web.

Existeixen diferents llenguatges de programació web, cadascun dels quals està enfocat a unes funcions en concret, i dels que convé destacar els següents (2):

HTML (Hypertext Markup Language): emprat per a descriure pàgines web.

CSS (Cascading Style Sheets): utilitzat per a descriure l'estètica dels elements HTML.

JavaScript: codi de programació web i de HTML per a crear pàgines interactives.

4. LA REGRESSIÓ LINEAL MÚLTIPLE

4.1. Introducció matemàtica de la regressió lineal múltiple

Les dades de les que parteix la regressió lineal múltiple són les $i = 1, \dots, n$ observacions de cada un dels $p+1$ conjunts de dades: $(x_1, x_2, \dots, x_{(p-1)}, x_p, y)$.

	x_1	x_2	...	x_{p-1}	x_p	y
1	x_{11}	x_{21}	...	$x_{(p-1)1}$	x_{p1}	y_1
2	x_{12}	x_{22}	...	$x_{(p-1)2}$	x_{p2}	y_2
...
n-1	$x_{(n-1)1}$	$x_{(n-1)2}$...	$x_{(n-1)(p-1)}$	$x_{(n-1)p}$	$y_{(n-1)}$
n	x_{n1}	x_{n2}	...	$x_{n(p-1)}$	x_{np}	y_n

Taula 4.1. Taula genèrica de dades.

El model de la regressió lineal múltiple és el següent:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i \quad (\text{Eq. 5.1})$$

$$i = 1, \dots, n \quad \varepsilon_i \sim N(0, \sigma^2)$$

On $\beta_0, \beta_1, \dots, \beta_p$ són els coeficients de regressió i ε_i és una seqüència aleatòria amb mitjana zero i variància σ^2 . Y és la resposta o variable dependent, mentre que X_1, X_2, \dots, X_p són les variables explicatives o independents.

Utilitzant notació matricial:

$$Y = X\beta + \varepsilon \quad (\text{Eq. 5.2})$$

Emprant el mètode dels mínims quadrats s'arriba a trobar l'estimador dels coeficients de la regressió:

$$\hat{\beta} = \min_{\beta_0, \beta_1, \dots, \beta_p} \sum_{i=1}^n \left(y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) \right)^2 \quad (\text{Eq. 5.3})$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (\text{Eq. 5.4})$$

Fer notar que l'accent circumflex en una variable indica que es tracta d'una estimació de la variable real en qüestió.

D'aquesta manera podem trobar per a uns valors $X = X_z$ els valors esperats de Y com:

$$\hat{Y} = X_z \hat{\beta} \quad (\text{Eq. 5.5})$$

Així doncs, es comprova que $\hat{\beta}$ és l'efecte estimat de la variable X_i en Y . És el canvi esperat a Y quan X_i augmenta en una unitat i el resta de variables romanen constants.

Els residus del model (el contingut estadístic que el model no explica) són:

$$\hat{\varepsilon} = Y - \hat{Y} \quad (\text{Eq. 5.6})$$

D'altra banda, cal destacar que la variància de l'estimat $\hat{\beta}$ és:

$$\text{var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} \quad (\text{Eq. 5.7})$$

I que un estimador no esbiaixat de σ^2 és:

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n - \dim(\beta)} = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n - (p+1)} \quad (\text{Eq. 5.8})$$

Convé ressaltar que és important d'evitar variables explicatives correlacionades entre si. Aquest fenomen de tenir les variables independents altament correlacionades és conegut com multicollinearitat. Aquesta situació s'ha d'evitar ja que augmenta la desviació estàndard dels coeficients de les variables correlacionades

L'interval de confiança del $100(1 - \alpha)\%$ de $\hat{\beta}$ és:

$$IC\ 100\ (1 - \alpha) = \hat{\beta} \pm t_{(1-\alpha/2), (n-(p+1))} \sqrt{\text{var}(\hat{\beta})} \quad (\text{Eq. 5.9})$$

On $t_{1-\alpha/2, (n-(p+1))}$ és el quantil $1 - \alpha/2$ de la distribució t de Student amb $n - (p + 1)$ graus de llibertat. Convé destacar que per a valors elevats de n i per a p molt més petit que n ($n \gg p$), és aplicable la simplificació $t_{(1-0.025), (n-(p+1))} = 1,96$.

Finalment, l'interval de predicció del $100(1 - \alpha)\%$ de \hat{Y} és:

$$IP\ 100(1 - \alpha)\% = \hat{Y} \pm t_{(1-\alpha/2), (n-(p+1))} \hat{\sigma} \sqrt{1 + X_z^T (X^T X)^{-1} X_z} \quad (\text{Eq. 5.11}) \quad (3)$$

4.2. La comprovació del model

En primer lloc, convé recordar que el model de la regressió lineal múltiple (Eq. 5.1) parteix de la suposició que $\varepsilon_i \sim N(0, \sigma^2)$ és una seqüència aleatòria amb mitjana zero i variància σ^2 . Així doncs, és important comprovar-ho abans d'usar el model obtingut.

Hi ha múltiples maneres de comprovar els supòsits del model utilitzant els residus o els residus estandarditzats (definites com la divisió dels residus entre un estimat de la seva variància). Els mètodes emprats per a la comprovació de l'anàlisi seran el següents:

La normalitat dels residus es comprova utilitzant gràfics qq (que exposen gràficament els quantils dels residus estandarditzats contra els quantils teòrics seguits per una normal). En cas de verificar-se el supòsit, els punts quedaran alineats en forma de línia recta. A la següent imatge es pot observar un gràfic qq d'un model on es verifiquen els supòsits.

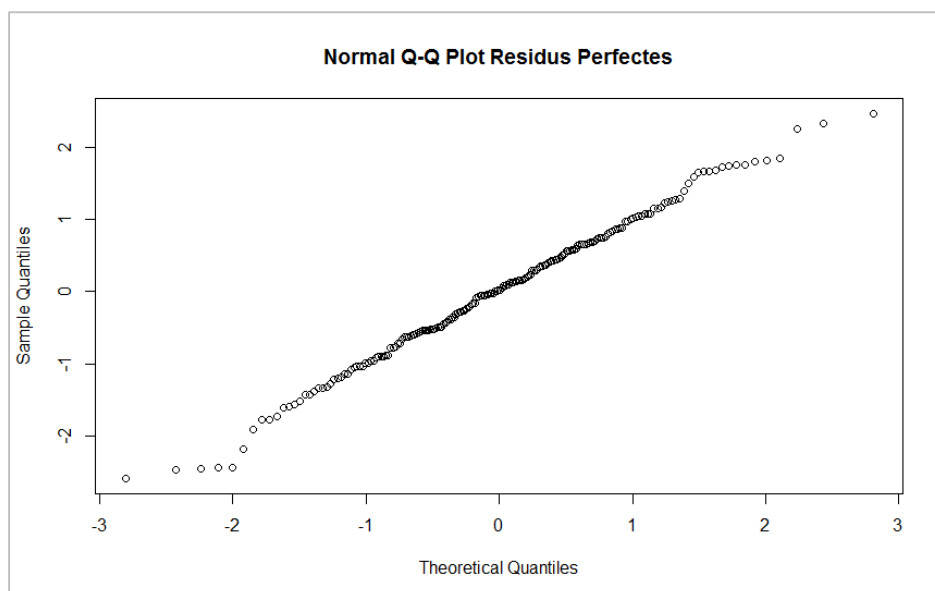


Figura 4.1. Exemple de gràfic qq en el qual es verifica la normalitat.

La homogeneïtat constant dels residus es comprova gràficament representant els residus estandarditzats contra els respectius valors estimats \hat{Y} pel model. En cas de verificar-se el supòsit, els residus tindran un patró aleatori, es situaran al voltant del zero i estaran ubicats ocupant una banda constant horitzontal.

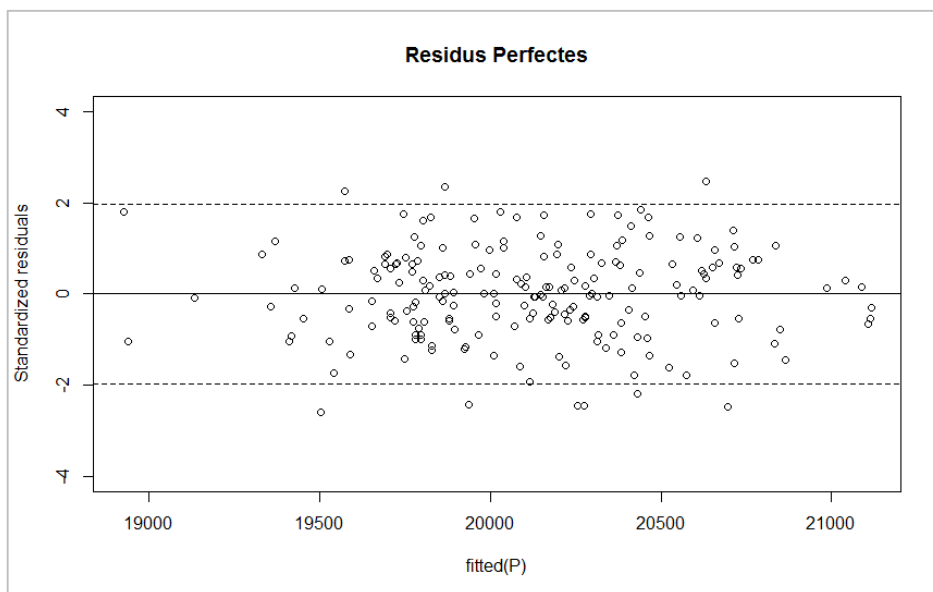


Figura 4.2. Exemple de verificació d'homogeneïtat constant dels residus.

L'efecte lineal de X_1, \dots, X_p es comprova representant gràficament els residus estandarditzats contra cada una de les variables explicatives. Altre cop s'espera un patró aleatori. De no ser així, es podria provar de aplicar alguna transformació a la resposta o d'afegir termes quadràtics o d'ordre superior al model.

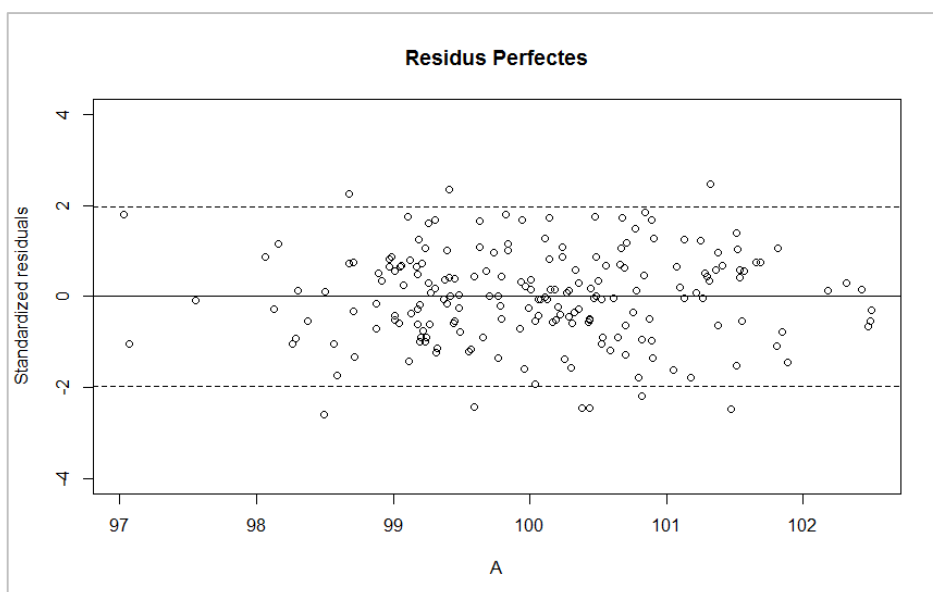


Figura 4.3. Exemple de verificació de l'efecte lineal d'una variable explicativa.

4.3. El mètode de l'eliminació progressiva (*Backward Selection*)

4.3.1. El mètode tradicional

Per definició, el p-valor indica la probabilitat d'obtenir un valor igual o més extrem a l'obtingut amb les dades, en el cas de ser certa la hipòtesi nul·la H_0 . És a dir, és el mínim nivell de significança estadística α (la probabilitat de rebutjar la hipòtesi nul·la sent aquesta certa) que conduiria a rebutjar la hipòtesi nul·la. Els p-valors són probabilitats associades als estadístics mostrals.

Per a la hipòtesi nul·la:

$$H_0: \beta_j = 0$$

I la hipòtesi alternativa:

$$H_1: \beta_j \neq 0$$

Es pot comprovar que l'estadístic mostral t segueix una distribució $t_{(n-(p+1))}$:

$$t = \frac{\hat{\beta}_j}{\sqrt{\text{var}(\hat{\beta}_j)}} \sim t_{(n-(p+1))}$$

I en conseqüència el p-valor:

$$p\text{-valor} = 2 \cdot P(T > |t|) = 2 \cdot (1 - P(T < |t|)) \quad (\text{Eq. 5.12})$$

on T és una variable aleatòria continua que segueix la distribució:

$$T \sim t_{(n-(p+1))}$$

Així doncs, es pot dur a terme l'eliminació progressiva de coeficients (*Backward Selection*), de la manera tradicional de la següent manera: es comença amb el model més gran (aquell que inclou totes les variables), i es remouen les variables que no són significants (aquelles en que no descartem la hipòtesi nul·la). Es descarten les variables d'una en una, i es continua fins que totes són significants.

De la mateixa manera, també es pot dur a terme el mètode de la introducció progressiva (*Forwards Selection*), consistent a començar el model amb exclusivament β_0 i anar ampliant a cada pas amb la variable més significativa, fins que no queden variables significatives fora del model. Així mateix també es pot emprar mètodes que combinen els dos anteriors.

4.3.2. El mètode modern

El mètode tradicional de selecció per eliminació progressiva té la avantatge de ser molt intuïtiu i fàcil d'entendre, però els programes de software moderns utilitzen un mètode més eficient i lleugerament diferent.

El criteri d'informació d'Akaike (AIC) d'un model M es defineix com:

$$AIC(M) = 2K - 2 \ln(L) \quad (\text{Eq 5.13}) \quad (4)$$

On K és el nombre de paràmetres del model i L la màxima versemblança d'aquest. Per a la regressió lineal múltiple, $K=p+2$.

L'AIC permet seleccionar el model òptim. Quan es dubta entre dos models, l'òptim serà aquell que tingui l'AIC més baix.

Així doncs, el mètode de modern d'eliminació progressiva parteix també del model més complex, el que inclou totes les variables. Altre cop es treuen les variables d'una en una, seleccionant el model amb l'AIC més baix, fins a trobar el mínim.

5. FILTRAT DE LES DADES

5.1. Dades d'interès

Per a tal de dur a terme l'anàlisi estadístic, és imprescindible de tractar les dades, eliminant aquelles que no són d'interès, així com reordenant les dades d'interès en un format més òptim per al seu posterior estudi.

L'ús de la regressió lineal múltiple requereix d'un vector amb les dades de la variable dependent o resposta (un vector amb les notes de l'assignatura a preveure) i un vector per a cada una de les variables independents o explicatives (un vector amb les notes respectives per a cada una de les assignatures que s'empraran per a preveure la nota de l'assignatura d'interès). Aquests vectors poden ser per a R o bé objectes diferents, o bé diverses columnes d'una mateixa matriu o taula.

Així doncs, es buscarà un algoritme que, donats els fitxers inicials amb les dades acadèmiques dels estudiants, retorni una matriu on cada fila representi un estudiant i cada columna inclogui la respectiva nota de cada una de les assignatures de primer curs del Grau en Enginyeria en Tecnologies Industrials, per a la primera convocatòria, excepte d'una columna que donarà la Nota d'Accés a la universitat.

5.2. L'algoritme del filtrat de les dades

L'algoritme emprat per a filtrar les dades funciona de la següent manera:

En primer lloc, crida els dos fitxers inicials de dades. Seguidament, es crea dos objectes del tipus *list*. El primer, anomenat F1, compost per les columnes 1, 2, 3, 5, 6 i 10 del fitxer *Dadestotesqualif1erfins2016.csv*. Aquestes columnes corresponen respectivament al codi del programa, el codi de l'expedient de l'alumne, el codi de l'assignatura, l'any i el quadrimestre de convocatòria, i la nota definitiva.

El segon objecte, anomenat F2, és compost per les columnes 2, 5 i 8 del fitxer *DadespersINDnombrespreins2016.csv*. Aquestes corresponen al codi de l'estudiant, l'any d'accés i la nota d'admissió a la universitat.

```

Dadestotesqualifierfins2016 <- read.csv2("C:/users/guill/Desktop/Dadestotesqualifierfins2016.csv")
DadespersINDnombrespreins2016 <- read.csv2("C:/users/guill/Desktop/DadespersINDnombrespreins2016.csv")

# F1 = [codi programa, codi expedient, codi upc, curs, quadrimestre, nota]
F1 = na.omit(cbind(Dadestotesqualifierfins2016[1:3], Dadestotesqualifierfins2016[5:6], Dadestotesqualifierfins2016[10]))
# F2 = [codi expedient, any acces, nota acces]
F2 = na.omit(cbind(DadespersINDnombrespreins2016[2], DadespersINDnombrespreins2016[5], DadespersINDnombrespreins2016[8]))

```

Figura 5.1. Crida dels fitxer inicials de dades i creació de F1 i F2.

A continuació, es procedeix a crear una matriu nul·la anomenada FF amb 1 fila i 17 columnes. Més tard s'afegiran a la matriu tantes files com estudiants s'hagin d'analitzar amb la regressió múltiple. Es considerarà a partir d'ara que una fila amb tot zeros és una fila buida. La primera columna de la matriu donarà el codi de l'estudiant, la segona la nota de la selectivitat, de la tercera a la setena inclouran la nota de les assignatures del primer quadrimestre, de la vuitena columna a la dotzena les notes del segon quadrimestre, i les últimes cinc columnes seran utilitzades per a assegurar-se que la nota que roman a la matriu és correspon a la primera convocatòria.

Seguidament es comença un bucle iteratiu. El bucle itera per a cada fila de F1. És a dir, cada iteració fa referència a un estudiant, una assignatura i una convocatòria en concret. La iteració no fa res si l'estudiant no cursa industrials (no té el codi 752 a la primera columna de F1) o si el codi de l'estudiant de la iteració no està també entre els codis dels estudiants de F2. D'aquesta manera es filtren tots els estudiants d'altres titulacions o dels que no tenim dades.

Pel contrari, de complir-se la condició anterior, es poden donar dos casos:

En primer lloc, que l'assignatura sigui del primer quadrimestre del grau. De ser així, la tercera columna de F1 tindrà un dels codis referents a una de les assignatures d'aquest quadrimestre. Cal que es compleixi aleshores que coincideixi l'any de la convocatòria amb l'any que l'estudiant va iniciar la universitat, i que el quadrimestre de la convocatòria sigui el primer. Aquesta condició es pot emprar ja que tots els estudiants estan obligats a cursar el primer quadrimestre totes les assignatures d'aquest.

L'algoritme comprova si l'estudiant té creada ja la seva fila a la matriu FF. De no ser així, s'omple la última fila de la matriu (que pel moment estarà buida). L'omple amb el codi de l'estudiant, els valors referents a la nota de l'assignatura respectiva de la iteració i amb el valor de la nota d'accés a la universitat, en diferents columnes. També es crea una nova darrera fila buida a la matriu FF. Si pel contrari l'estudiant ja tenia creada la fila, tant sols es guarda la nota de l'assignatura.

En segon lloc, es pot donar el cas que l'assignatura sigui del segon quadrimestre. F1 tindrà llavors a la tercera columna un dels codis respectius a les corresponents assignatures. L'algoritme funciona llavors similar al cas del primer quadrimestre, però amb la diferencia que quan es guarda la nota d'un estudiant d'una assignatura, també es guarda en una altra columna auxiliar (una de les cinc últimes columnes de FF, una per a cada assignatura) el valor corresponent a un comptador $sum = 2 \cdot any\ convocatòria + quadrimestre$ (1 o 2). Es comprova que el valor de *sum* queda ordenat cronològicament. Així doncs, abans d'omplir la columna de la nota, l'algoritme comprova que o bé la columna del comptador *sum* corresponent a aquesta estigui buit, o bé que el d'aquesta convocatòria sigui més petit que el valor que ocupa el comptador en aquell moment.

```
FF = matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0), ncol=17)
j=0
for (i in c(1:length(F1[,1]))) {
  if (F1[i,1]==752 & F1[i,2] %in% F2[,1]) {
    mtch = match(F1[i,2], F2[,1])
    if (F1[i,3] == 240011 | F1[i,3] == 240012 | F1[i,3] == 240013 | F1[i,3] == 240014 | F1[i,3] == 240015) {
      if (F2[mtch, 2] == F1[i,4] & F1[i,5]==1) {
        if (F1[i,2] %in% FF[,1]==FALSE){
          FF[j,1]=F1[i,2]
          FF[j,2]= F2[mtch, 3]
          j=j+1
          FF = rbind(FF, c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))
          mtch2 = match(F1[i,2], FF[,1])
          if (F1[i,3] == 240011) {FF[mtch2,3]=F1[i,6]}
          if (F1[i,3] == 240012) {FF[mtch2,4]=F1[i,6]}
          if (F1[i,3] == 240013) {FF[mtch2,5]=F1[i,6]}
          if (F1[i,3] == 240014) {FF[mtch2,6]=F1[i,6]}
          if (F1[i,3] == 240015) {FF[mtch2,7]=F1[i,6]}
        }
      }
    }
    if (F1[i,3] == 240021 | F1[i,3] == 240022 | F1[i,3] == 240023 | F1[i,3] == 240024 | F1[i,3] == 240025){
      if (F1[i,2] %in% FF[,1]==FALSE){
        FF[j,1]=F1[i,2]
        FF[j,2]= F2[mtch, 3]
        col = F1[i,3] - 240020 + 12
        sum = (2*F1[i,4])+F1[i,5]
        FF[j,col] = sum
        if (F1[i,3] == 240021) {FF[j,8]=F1[i,6]}
        if (F1[i,3] == 240022) {FF[j,9]=F1[i,6]}
        if (F1[i,3] == 240023) {FF[j,10]=F1[i,6]}
        if (F1[i,3] == 240024) {FF[j,11]=F1[i,6]}
        if (F1[i,3] == 240025) {FF[j,12]=F1[i,6]}
        j=j+1
        FF = rbind(FF, c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))
      }
    }
    else {
      col = F1[i,3] - 240020 + 12
      mtch2 = match(F1[i,2], FF[,1])
      sum = (2*F1[i,4])+F1[i,5]
      if (sum < FF[mtch2,col] | FF[mtch2,col]==0){
        FF[mtch2,col]=sum
        if (F1[i,3] == 240021) {FF[mtch2,8]=F1[i,6]}
        if (F1[i,3] == 240022) {FF[mtch2,9]=F1[i,6]}
        if (F1[i,3] == 240023) {FF[mtch2,10]=F1[i,6]}
        if (F1[i,3] == 240024) {FF[mtch2,11]=F1[i,6]}
        if (F1[i,3] == 240025) {FF[mtch2,12]=F1[i,6]}
      }
    }
  }
}
```

Figura 5.2. Creació de la matriu FF

Finalment, es procedeix a crear la matriu Dades. De la mateixa manera que amb la matriu FF, es procedeix en primer lloc en crear la matriu buida, amb una fila i 12 columnes. Les columnes seran corresponents a les primeres 12 columnes de la matriu FF.

El bucle iteratiu, comprova, que per a cada fila de FF, els comptadors *sum* siguin diferents de zero. De no ser així, significa que l'estudiant encara no ha cursat aquella assignatura. Com que el nombre de dades del que es disposa és molt elevat, es prescindeix, sense perdre informació estadística rellevant, dels estudiants que no han cursat totes les assignatures del segon quadrimestre i aquests no s'inclouen a la matriu Dades. A més, es comprova que la nota d'accés a la universitat sigui superior a 9. D'aquesta manera s'eliminen casos puntuals d'estudiants que no provenien del batxillerat i que, conseqüentment, se'ls requeria menor nota per a entrar al grau.

Per acabar, es borra la última fila de la matriu ja que és buida com a conseqüència de l'algoritme, i es posen nom a les columnes per a facilitar la seva interpretació.

```

Dades = matrix(c(0,0,0,0,0,0,0,0,0,0,0,0), ncol=12)
j=1
for (i in c(1:length(FF[,1]))) {
  if (FF[i,13] !=0 & FF[i,14] !=0 & FF[i,15] !=0 & FF[i,16] !=0 & FF[i,17] !=0 & FF[i,2]>=9) {
    Dades[j,]=FF[i,1:12]
    Dades = rbind(Dades, c(0,0,0,0,0,0,0,0,0,0,0,0))
    j=j+1
  }
}

Dades = Dades[1:length(Dades[,1])-1,]
colnames(Dades)=c("Codi Est.", "N.Acces", "Algebra", "CalculI", "Mec.", "Quim.I", "Infor.", "Geom.",
                  "CalculII", "Termo.", "Quim.II", "Expr.")
Dades

```

Figura 5.3. Creació de la matriu Dades

5.3. L'anàlisi en R de la regressió lineal múltiple

5.3.1. Anàlisi de correlacions entre les variables explicatives

És convenient recordar (del punt 4.3.1) que cal comprovar que no hi ha correlacions entre les variables explicatives d'un model de regressió lineal múltiple per tal d'evitar que es doni el fenomen de la col·linealitat i les seves conseqüències.

Així doncs, per a comprovar que no hi ha correlacions, en primer lloc, s'importa la llibreria *car* a R. La importació d'una llibreria permet utilitzar noves funcions que sense aquestes llibreries

no es podria. D'aquesta manera, usem la funció *scatterplotMatrix* de la llibreria *car* per a dibuixar una matriu de gràfics. El *scatterplot matrix* (matriu de diagrames de dispersió) d'unes variables dóna els diagrames de dispersió de cada una de les variables respecte el resta d'aquestes. D'aquesta manera es pot observar visualment l'existència de correlacions entre variables.

A la següent figura observem la matriu de diagrames de dispersió de les variables explicatives dels nostres models, amb els respectius *boxplots* (diagrames de caixa) a la seva diagonal.

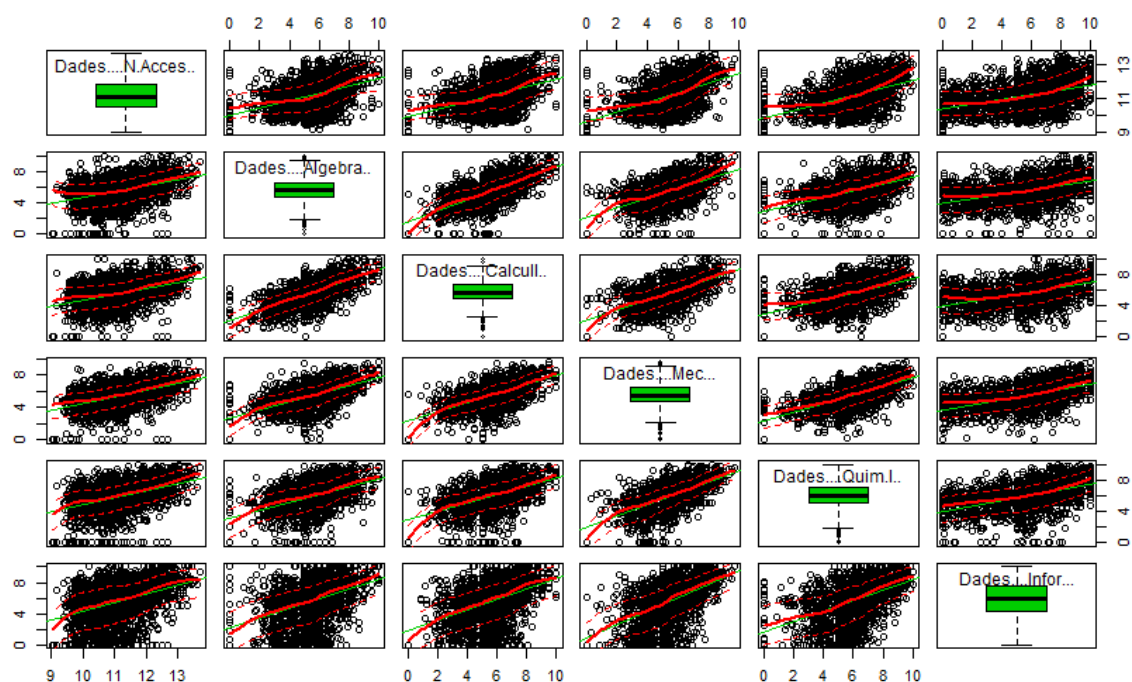


Figura 5.4. ScatterplotMatrix i boxplots de les variables explicatives.

S'observa que, *a priori*, no semblen haver-hi fortes correlacions entre variables, i per tant no es donarà el fenomen de col·linealitat en la regressió lineal múltiple. Això mateix es comprova calculant el coeficient de correlació de Pearson (r) d'algunes parelles de variables. Aquest coeficient dóna una mesura de la relació lineal entre dues variables. El coeficient varia entre 1 (correlació positiva) i -1 (correlació negativa), i dóna 0 en cas de no existir correlació.

Així doncs, es comprova el coeficient de correlació d'aquelles parelles de variables que en el *scatterplot matrix* semblen tenir més relació lineal són els següents:

$$r_{\text{Àlgebra-Càlcul I}} = 0,6647$$

$$r_{\text{Nota Accés-Mec. Fonamental}} = 0,4828$$

$$r_{Mec. Fonamental-Química I} = 0,6140$$

S'observa com els tres valors estan lluny de 1, i per tant, reforcen a concloure que no hi ha relació lineal entre les variables.

5.3.2. Realització dels models inicials de regressió lineal múltiple

Un cop comprovat que no hi ha elevades correlacions entre les variables explicatives del model, es procedeix a trobar els models de la regressió lineal múltiple de les notes del segon quadrimestre. Els models inicials amb els que es treballarà són aquelles que preveuen cada una de les notes del segon quadrimestre usant tot el conjunt de variables explicatives: totes les notes del primer quadrimestre i la nota d'accés a la universitat.

Així doncs, aplicant l'equació (Eq. 5.1) a aquests casos particulars arribem a la següent expressió:

$$Nota a preveure = C_0 + C_1 \cdot Nota Accés + C_2 \cdot N. Àlg. + C_3 \cdot N. Càl. I + C_4 \cdot N. Mec. Fon. + C_5 \cdot N. Quím. I + C_6 \cdot N. Fonam. Inf \quad (Eq. 5.14)$$

5.3.3. Comprovació dels supòsits dels models

L'objectiu final de l'anàlisi en R de les dades és reduir els models inicials fins a trobar es models òptims. El mètode que s'usarà per a la optimització dels models és l'eliminació progressiva. Tanmateix, per a poder usar aquest mètode és necessari de verificar prèviament les hipòtesis dels models inicials.

a) Normalitat dels residus

Convé recordar que la verificació de la normalitat dels residus es realitza visualment mitjançant els gràfics qq. A la següent figura es poden observar els gràfics obtinguts amb els residus dels models inicials. Es comprova que els punts obtinguts per els models segueixen en gran part una línia recta. Tot i no ser perfectes, s'adeqüen suficientment bé com per a continuar.

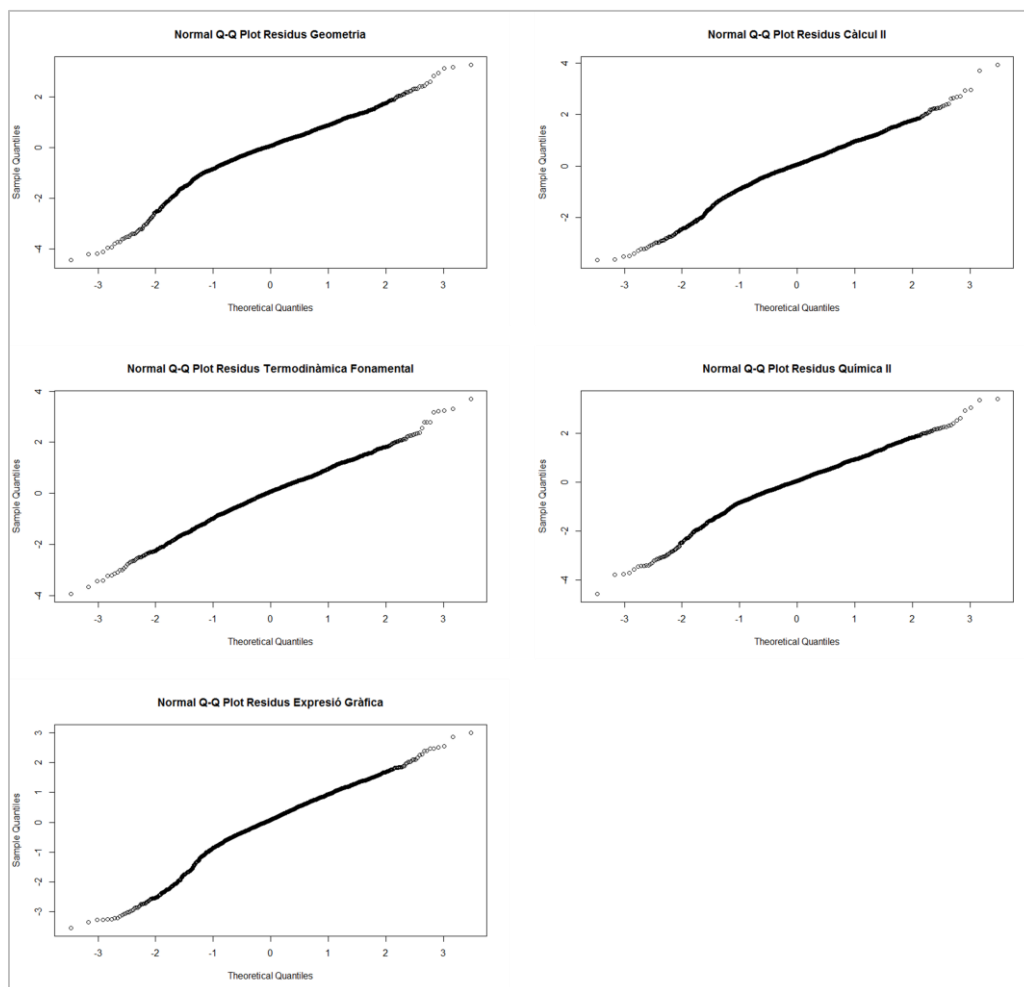


Figura 5.5. Gràfics qq dels residus dels models inicials.

b) Homogeneïtat constant dels residus

Cal recordar que per a comprovar la homogeneïtat constant dels residus es representa gràficament els residus estandarditzats contra els valors estimats \hat{Y} pel model. En cas de verificar-se el supòsit, els residus tindran un patró aleatori, es situaran al voltant del zero i estaran ubicats ocupant una banda constant horitzontal.

Així doncs, es comprova a la següent figura que per a tots els models es pot verificar la homogeneïtat constant dels residus.

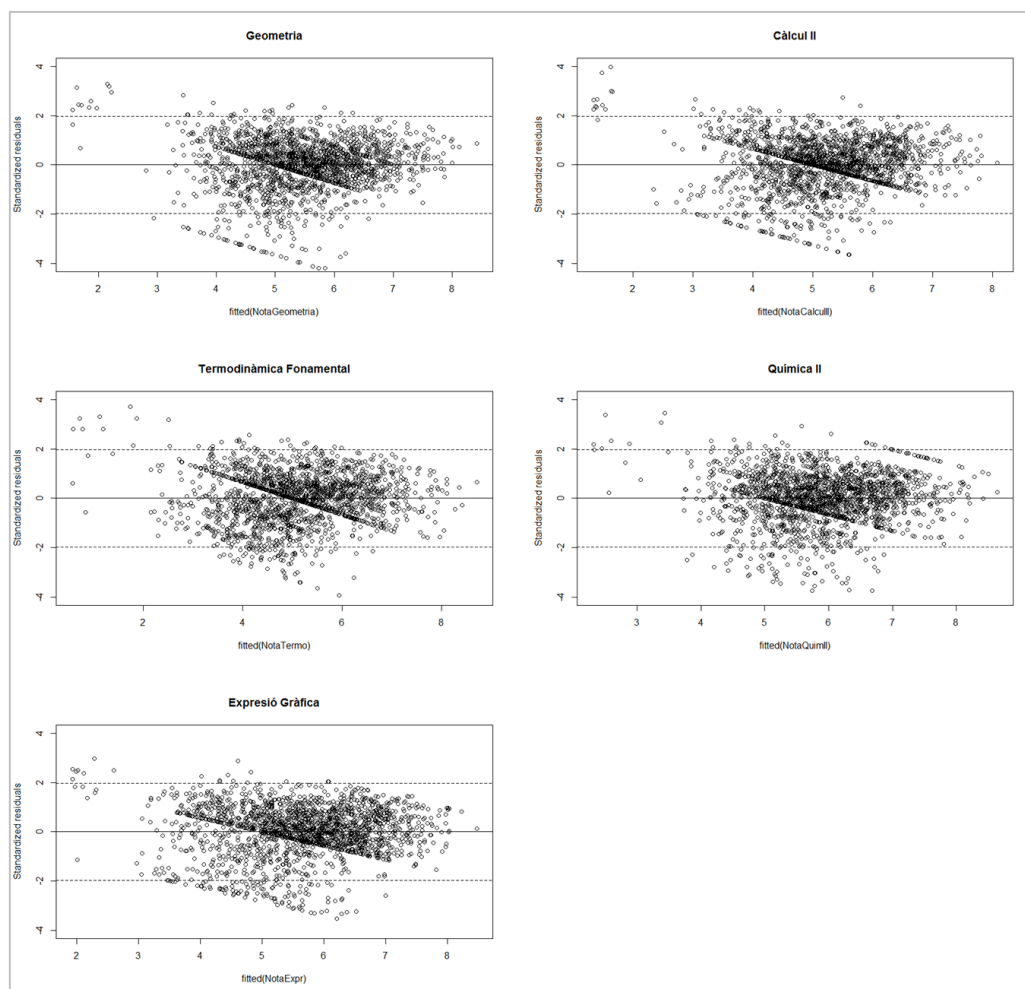


Figura 5.6. Gràfics del residus estandarditzats contra els valors estimats pel model.

c) Efecte lineal de les variables explicatives

Es comprova observant, per a cada un dels models, els gràfics dels residus estandarditzats contra cada una de les variables explicatives. D'aquesta manera, es verifica l'efecte lineal quan en el gràfic s'observa un patró aleatori. Les següents figures contenen els gràfics de totes les variables explicatives per a cada model respectivament.

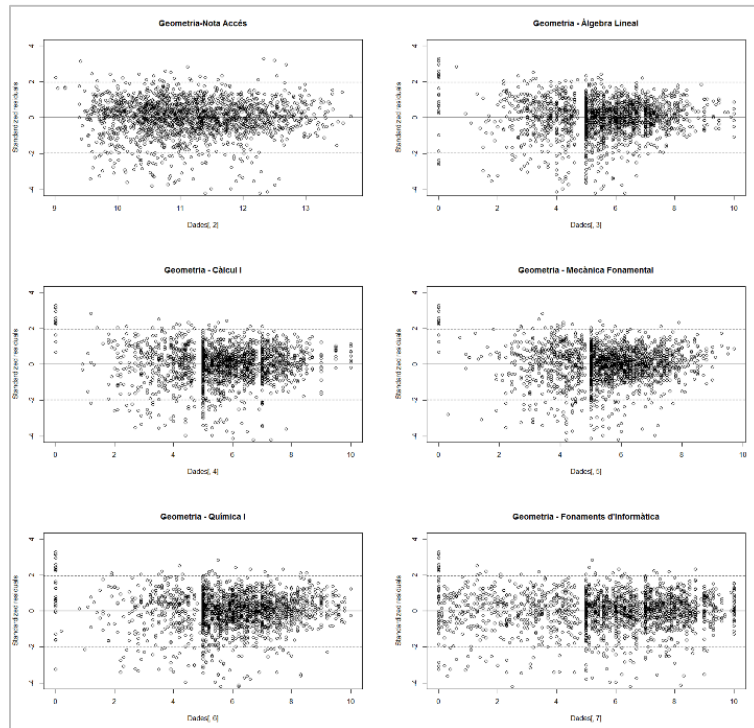


Figura 5.7. Gràfics de les variables explicatives pel model de Geometria.

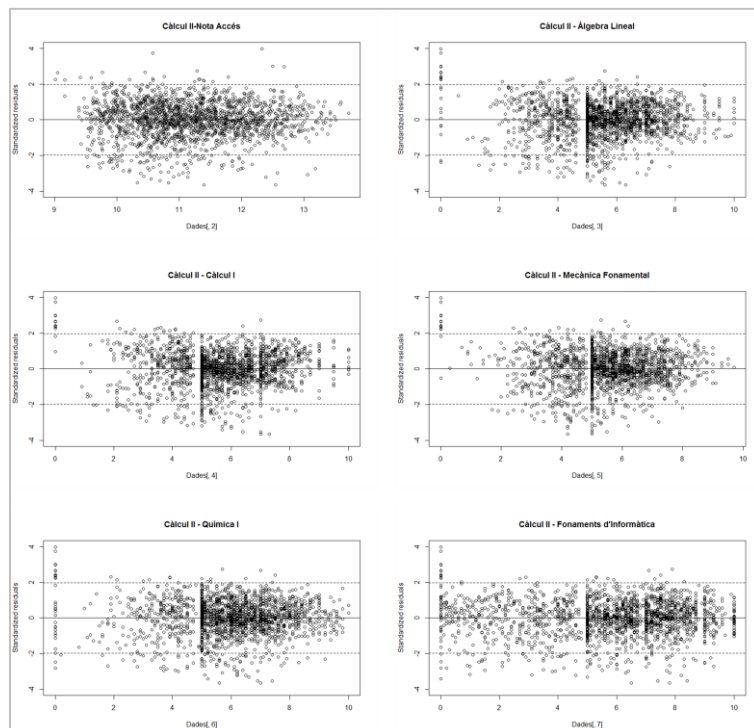


Figura 5.8. Gràfics de les variables explicatives pel model de Càlcul II.

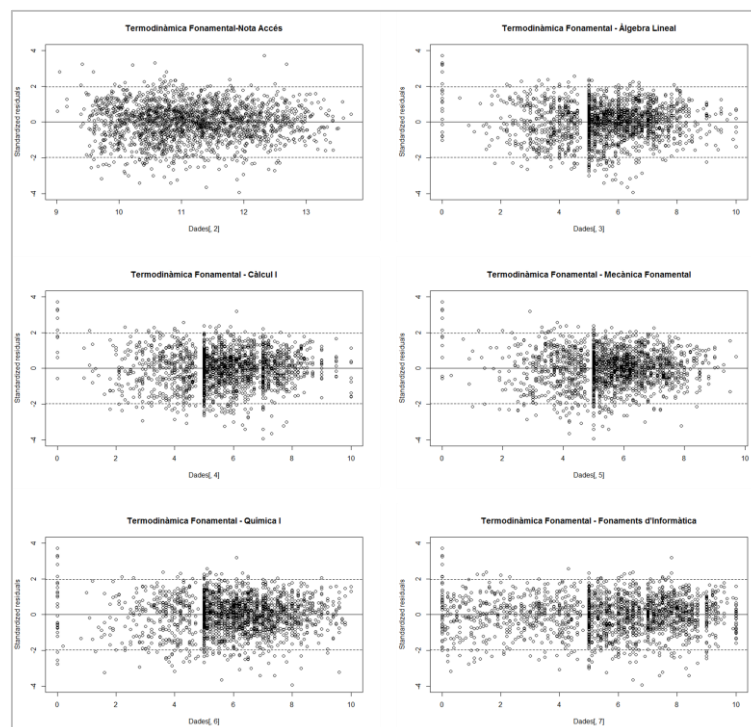


Figura 5.9. Gràfics de les variables explicatives pel model de Termodinàmica.

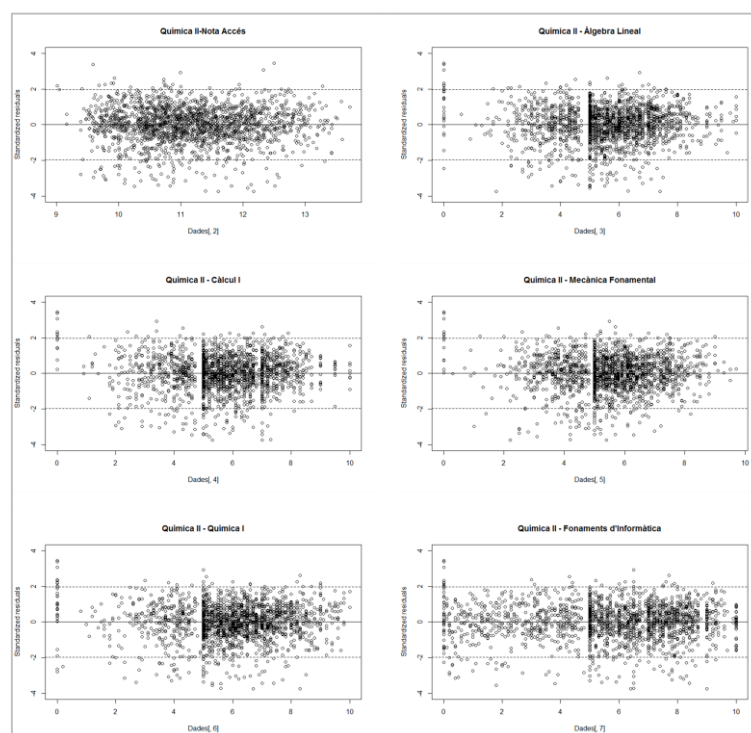


Figura 5.10. Gràfics de les variables explicatives pel model de Química II.

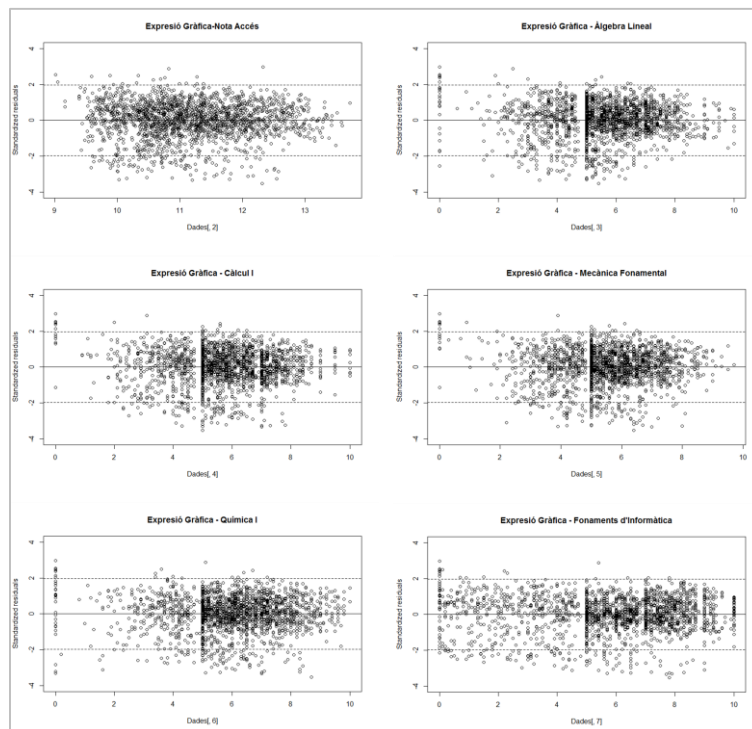


Figura 5.11. Gràfics de les variables explicatives pel model d'Expressió Gràfica.

5.3.4. Realització dels models finals de la regressió lineal múltiple (*Backward Selection*)

Un cop verificats els residus, es procedeix a trobar els models finals. Aquests models seran similars als inicials, però prescindint d'aquelles variables explicatives que no siguin necessàries pel model.

$$\begin{aligned} \text{Nota a preveure} = & C'_0 + C'_1 \cdot \text{Nota Accés} + C'_2 \cdot N. \text{Alg.} + C'_3 \cdot N. \text{Càl. I} + \\ & + C'_4 \cdot N. \text{Mec. Fon.} + C'_5 \cdot N. \text{Quím. I} + C'_6 \cdot N. \text{Fonam. Inf.} \end{aligned} \quad (\text{Eq. 5.15})$$

On C'_i serà igual a zero si és que la variable explicativa i no forma part del model.

Es procedeix doncs, a trobar el model final per a cada una de les assignatures del segon quadrimestre:

a) Geometria

En primer lloc, es calcula el AIC tant del model inicial com del model inicial traient cada una de les variables respectivament.

Nota Geometria	AIC del model sense la variable	p-valor: $C'i = 0$
Model Actual	1234,0	
Constant	1232,0	0,879770
Nota Accés	1250,5	$1,80 \cdot 10^{-5}$
Nota Àlgebra	1246,5	0,000146
Nota Càlcul I	1301,3	$<2 \cdot 10^{-16}$
Nota Mèc. Fon.	1273,1	$1,55 \cdot 10^{-10}$
Nota Quím. I	1235,5	0,060462
Nota Fonam. Inf.	1239,7	0,005500

Taula 5.1. AIC del model inicial de Geometria i del model sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$.

S'observa que l'AIC més baix s'aconsegueix retirant el terme independent del model. Així doncs, a la següent figura es mostren els AIC del model sense constant (ja mostrat també a la taula anterior) i els AIC dels models que a més a més de la constant, se'ls ha tret alguna altra variable:

Nota Geometria	AIC del model sense la variable	p-valor: $C'i = 0$
Model Actual	1232,0	
Nota Accés	1378,6	$<2 \cdot 10^{-16}$
Nota Àlgebra	1244,5	0,000145
Nota Càlcul I	1299,3	$<2 \cdot 10^{-16}$
Nota Mèc. Fon.	1271,5	$1,27 \cdot 10^{-10}$
Nota Quím. I	1233,6	0,058168
Nota Fonam. Inf.	1238,0	0,004760

Taula 5.2. AIC del model actual de Geometria i del model sense cada una de les variables, i el p-valor dels coeficients del model actual $C'i = 0$.

S'observa que altre cop l'AIC mínim s'aconsegueix amb el model que conte totes les variables menys la constant. Així doncs, ja no es retiren més variables i en conseqüència, aquest és el model definitiu per a Geometria.

b) Càlcul II

Com en l'anterior cas, en primer lloc es calcula l'AIC tant del model inicial com del model traient cada una de les variables respectivament.

Nota Càlcul II	AIC del model sense la variable	p-valor: $C'i = 0$
Model Actual	1694,2	
Constant	1693,7	0,2202
Nota Accés	1695,6	0,0661
Nota Àlgebra	1734,2	$1,01 \cdot 10^{-10}$
Nota Càlcul I	1782,2	$<2 \cdot 10^{-16}$
Nota Mèc. Fon.	1716,2	$1,03 \cdot 10^{-06}$
Nota Quím. I	1693,2	0,3065
Nota Fonam. Inf.	1698,3	0,0137

Taula 5.3. AIC del model inicial de Càlcul II i l'AIC d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$.

L' AIC més baix s'aconsegueix retirant el terme respectiu a la Nota de Química I. Tanmateix, a la següent figura es mostren els AIC del model sense aquest coeficient (ja mostrat també a la taula anterior) i els AIC dels models que a més a més se'ls ha tret alguna altra variable:

Nota Càlcul II	AIC del model sense la variable	p-valor: $C'i = 0$
Model Actual	1693,2	
Constant	1692,6	0,2489
Nota Accés	1695,4	0,0425
Nota Àlgebra	1734,5	$5,18 \cdot 10^{-11}$
Nota Càlcul I	1783,3	$<2 \cdot 10^{-16}$
Nota Mèc. Fon.	1721,8	$3,46 \cdot 10^{-08}$
Nota Fonam. Inf.	1696,7	0,0203

Taula 5.4. AIC del model actual de Càlcul II i del model sense cada una de les variables, i el p-valor dels coeficients del model actual $C'i = 0$.

S'observa que cal treure la constant del model. Així doncs es procedeix a tornar a observar si retirant alguna altra variable es pot reduir més l'AIC.

Nota Càlcul II	AIC del model sense la variable	p-valor: $C'i = 0$
Model Actual	1692,6	
Nota Accés	1779,9	$<2 \cdot 10^{-16}$
Nota Àlgebra	1733,5	$5,93 \cdot 10^{-11}$
Nota Càlcul I	1782,6	$<2 \cdot 10^{-16}$
Nota Mèc. Fon.	1720,1	$5,76 \cdot 10^{-08}$
Nota Fonam. Inf.	1697,0	0,0117

Taula 5.5. AIC del model actual de Càlcul II i del model sense cada una de les variables, i el p-valor dels coeficients del model actual $C'i = 0$.

Finalment s'observa que el model amb el AIC més baix és el que conté totes les variables excepte la constant i la nota de química I. Aquest serà doncs, el model definitiu per Càlcul II.

c) Termodinàmica Fonamental

Altres cop es calcula en primer lloc l'AIC tant del model inicial com del model traient cada una de les variables respectivament.

Nota Term. Fonamental	AIC del model sense la variable	p-valor: $C'i = 0$
Model Actual	1625,8	
Constant	1656,6	$1,07 \cdot 10^{-8}$
Nota Accés	1680,9	$4,65 \cdot 10^{-14}$
Nota Àlgebra	1769,0	$<2 \cdot 10^{-16}$
Nota Càlcul I	1645,0	$4,32 \cdot 10^{-06}$
Nota Mèc. Fon.	1636,2	0,000427
Nota Quím. I	1627,6	0,052265
Nota Fonam. Inf.	1623,9	0,756210

Taula 5.6. AIC del model inicial de Termodinàmica Fonamental i l'AIC d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$.

S'observa que el AIC més baix s'obté traient la nota de Fonaments d'Informàtica del model. D'aquesta manera, es mostra a continuació l'AIC del model inicial sense aquest coeficient i els AIC dels models que a més se'ls ha tret alguna altra variable:

Nota Term. Fonamental	AIC del model sense la variable	p-valor: $C'i = 0$
Model Actual	1623,9	
Constant	1654,8	$9,98 \cdot 10^{-9}$
Nota Accés	1679,3	$3,79 \cdot 10^{-14}$
Nota Àlgebra	1767,2	$< 2 \cdot 10^{-16}$
Nota Càlcul I	1643,0	$4,5 \cdot 10^{-6}$
Nota Mèc. Fon.	1634,4	0,000419
Nota Quím. I	1625,6	0,055256

Taula 5.7. AIC del model actual de Termodinàmica Fonamental i d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$.

S'observa que l'AIC mínim s'aconsegueix amb el model que conte totes les variables menys la constant. Així doncs, ja no es retiren més variables i en conseqüència aquest és el model definitiu per a Termodinàmica Fonamental.

d) Química II

De la mateix manera que en els anteriors casos, calcula en primer lloc l'AIC tant del model inicial com del model traient cada una de les variables respectivament.

Nota Química II	AIC del model sense la variable	p-valor: $C'i = 0$
Model Actual	1637,9	
Constant	1637,4	0,211
Nota Accés	1683,1	$6,94 \cdot 10^{-12}$
Nota Àlgebra	1673,7	$8,53 \cdot 10^{-10}$
Nota Càlcul I	1663,5	$1,52 \cdot 10^{-07}$
Nota Mèc. Fon.	1636,2	0,561
Nota Quím. I	1653,9	$2,25 \cdot 10^{-05}$
Nota Fonam. Inf.	1638,1	0,138

Taula 5.8. AIC del model inicial de Química II i l'AIC d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$.

S'observa que per a optimitzar el model cal treure la nota de Mecànica Fonamental del model. A continuació es continua observant si traient més variables s'aconsegueixen AIC més baixos.

Nota Química II	AIC del model sense la variable	p-valor: $C'i = 0$
Model Actual	1636,2	
Constant	1635,9	0,194
Nota Accés	1684,1	$1,79 \cdot 10^{-12}$
Nota Àlgebra	1674,0	$3,00 \cdot 10^{-10}$
Nota Càlcul I	1665,2	$2,81 \cdot 10^{-08}$
Nota Quím. I	1656,2	$2,84 \cdot 10^{-06}$
Nota Fonam. Inf.	1636,8	0,105

Taula 5.9. AIC del model actual de Química II i d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$.

A l'anterior taula s'observa que cal retirar la constant del model. Després, es torna a comprovar si cal treure més variables.

Nota Química II	AIC del model sense la variable	p-valor: $C'i = 0$
Model Actual	1635,9	
Nota Accés	1934,7	$<2 \cdot 10^{-16}$
Nota Àlgebra	1674,1	$2,45 \cdot 10^{-10}$
Nota Càlcul I	1665,0	$2,56 \cdot 10^{-08}$
Nota Quím. I	1657,4	$1,28 \cdot 10^{-06}$
Nota Fonam. Inf.	1637,3	0,0661

Taula 5.10. AIC del model actual de Química II i d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$.

Finalment, s'observa que el sense constant ni a nota de Mecànica Fonamental és l'òptim, ja que aquest model és el que té el AIC més baix.

e) Expressió Gràfica

En últim lloc, es procedeix a optimitzar el model per a Expressió Gràfica. Es mostra a continuació l'AIC tant del model inicial com del model traient cada una de les variables respectivament.

Nota Expressió Gràfica	AIC del model sense la variable	p-valor: $C'i = 0$
Model Actual	2239,6	
Constant	2241,1	0,063161
Nota Accés	2241,4	0,050754
Nota Àlgebra	2246,8	0,002421
Nota Càlcul I	2252,1	0,000147
Nota Mèc. Fon.	2252,7	0,000105
Nota Quím. I	2238,7	0,292019
Nota Fonam. Inf.	2387,9	$<2 \cdot 10^{-16}$

Taula 5.11. AIC del model inicial d'Expressió Gràfica i l'AIC d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$.

S'observa que l'AIC mínim s'obté si es retira del model la nota de Química I. A continuació es tornen a calcular els AIC tant del model sense la nota de Química I com d'aquest model sense cada una de les variables respectivament.

Nota Expressió Gràfica	AIC del model sense la variable	p-valor: $C'i = 0$
Model Actual	2238,7	
Constant	2240,5	0,052143
Nota Accés	2240,0	0,070665
Nota Àlgebra	2245,5	0,003084
Nota Càlcul I	2250,6	0,000198
Nota Mèc. Fon.	2250,8	0,000184
Nota Fonam. Inf.	2387,3	$<2 \cdot 10^{-16}$

Taula 5.12. AIC del model actual d'Expressió Gràfica i d'aquest sense cada una de les variables, i el p-valor dels coeficients del model $C'i = 0$.

S'observa que l'AIC mínim s'aconsegueix amb el model que conte totes les variables menys la nota de Química I. Així doncs, ja no es retiren més variables. Aquest és el model definitiu per a Expressió Gràfica.

5.4. El models definitius

En primer lloc, convé recordar que el valor estimat del coeficients d'un model de regressió lineal múltiple és l'efecte estimat de la variable respectiva al coeficient a la resposta (en aquest cas la nota de l'assignatura a preveure). És el canvi esperat a Y quan X_i augmenta en una unitat i el resta de variables romanen constants. Així doncs, és important de presentar els coeficients obtinguts en cada model.

Nota Geometria	Coefficient
Nota Accés	0,17365
Nota Àlgebra	0,10130
Nota Càlcul I	0,23829
Nota Mèc. Fon.	0,20059
Nota Quím. I	0,04501
Nota Fonam. Inf.	0,04331

Taula 5.13. Coeficients del model final de Geometria.

Nota Càlcul II	Coefficient
Nota Accés	0,14480
Nota Àlgebra	0,19654
Nota Càlcul I	0,30916
Nota Mèc. Fon.	0,17936
Nota Fonam. Inf.	-0,04275

Taula 5.14. Coeficients del model final de Càlcul II.

Nota Term. Fonamental	Coefficient
Constant	-2,56291
Nota Accés	0,34919
Nota Àlgebra	0,36007
Nota Càlcul I	0,14405
Nota Mèc. Fon.	0,11942
Nota Quím. I	0,04968

Taula 5.15. Coeficients del model final de Termodinàmica Fonamental.

Nota Química II	Coefficient
Nota Accés	0,26822
Nota Àlgebra	0,18557
Nota Càlcul I	0,17113
Nota Quím. I	0,12067
Nota Fonam. Inf.	0,03063

Taula 5.16. Coeficients del model final de Química II.

Nota Expressió Gràfica	Coefficient
Constant	1,01971
Nota Accés	0,09657
Nota Àlgebra	0,10178
Nota Càlcul I	0,13681
Nota Mèc. Fon.	0,14271
Nota Fonam. Inf.	0,24617

Taula 5.17. Coeficients del model final d'Expressió Gràfica.

6. L'APLICACIÓ WEB

6.1. La interfície d'usuari

L'HTML i el CSS permeten descriure pàgines web i la seva estètica. Tot i que l'HTML és un codi que funciona per etiquetes, els navegadors web (Internet Explorer, Google Chrome, Firefox, Safari, etc.) no mostren aquestes, sinó que les utilitzen per a mostrar el document. A la següent imatge es mostra el que el navegador web ensenya per pantalla en referència al codi de programació de l'aplicació web pel càlcul estadístic de les notes.

Ompli el següent formulari amb els resultats acadèmics corresponents i premi el botó Continuar. Usar un punt com a separador decimal.

Àlgebra Lineal (240011):	<input type="text"/>
Càlcul I (240012):	<input type="text"/>
Mecànica Fonamental (240013):	<input type="text"/>
Química I (240014):	<input type="text"/>
Fonaments d'Informàtica (240015):	<input type="text"/>
Nota d'Accés a la Universitat:	<input type="text"/>

Assignatura a consultar:

Geometria (240021)

Aquesta és una aplicació de caràcter acadèmic.
Per a qualsevol dubte contacteu amb guillembalcells@hotmail.com

Figura 6.1. Interfície d'usuari de l'aplicació.

6.2. El JavaScript

JavaScript permet programar el comportament de les pàgines web i aconseguir que aquestes siguin interactives. S'ensenya a continuació el codi JavaScript de l'aplicació web. Concretament, es mostra en primer lloc la *Funcio1*, que es crida quan l'usuari prem el botó que a l'anterior imatge hi té escrit "Continuar".

La primera tasca que realitza la *Funcio1* és guardar-se els valors referents als resultats acadèmics que l'usuari ha introduït i a l'assignatura la nota de la qual vol preveure. Convé recordar que al primer quadrimestre del Grau en Enginyeria en Tecnologies Industrials, per aquells estudiants que provenen de cursar un Batxillerat i les Proves d'Accés a la Universitat, és obligatori de cursar totes cinc assignatures.

Si algun valor no ha estat introduït, algun valor no és de caràcter numèric o s'ha introduït algun valor inferior al zero o superior al deu, aleshores l'aplicació farà un avís, informant a l'usuari que alguna nota es troba fora de rang.

En qualsevol altra cas, l'aplicació cridarà la *Funcio2*, encarregada de realitzar els càlculs estadístics pertinents i de mostrar el resultat per pantalla. També canviarà el text del botó de "Continuar" a "Actualitzar" i s'actualitzarà el text informatiu de la part superior de l'aplicació web.

```
function Funcio1() {
    var text;

    nota1 = document.getElementById("nota1").value;
    nota2 = document.getElementById("nota2").value;
    nota3 = document.getElementById("nota3").value;
    nota4 = document.getElementById("nota4").value;
    nota5 = document.getElementById("nota5").value;
    notaA = document.getElementById("notaA").value;
    assignatura = document.getElementById("assignatura").value;

    // si x es Not a Number o mes petit que 0 (9) o mes gran 10 (14)
    if (isNaN(nota1) || nota1 < 0 || nota1 > 10 || isNaN(nota2) || nota2 < 0 || nota2 > 10 || isNaN(nota3) ||
        nota3 < 0 || nota3 > 10 || isNaN(nota4) || nota4 < 0 || nota4 > 10 || isNaN(nota5) || nota5 < 0 || nota5 > 10 ||
        isNaN(notaA) || notaA < 9 || notaA > 14 || nota1==" || nota2==" || nota3==" || nota4==" || nota5==" ) {
        text = "Alguna nota es troba fora de rang. Revisi el formulari i premi el botó Continuar.";
        document.getElementById("tpunt").innerHTML = "Nota: usar un punt com a separador decimal";
        document.getElementById("tresultat").innerHTML = "";
    } else {
        text = "Premi Actualitzar per a veure la seva nota prevista de l'assignatura que seleccioni.";
        document.getElementById("textb1").innerHTML = "Actualitzar";
        document.getElementById("tpunt").innerHTML = "";
        Funcio2();
    }
    document.getElementById("info").innerHTML = text;
}
```

Figura 6.2. Codi de la Funcio1.

APP pel càlcul estadístic de les notes

Alguna nota es troba fora de rang. Revisi el formulari i premi el botó Continuar.

Àlgebra Lineal (240011):	6
Càlcul I (240012):	5.5
Mecànica Fonamental (240013):	5.2
Química I (240014):	10.2
Fonaments d'Informàtica (240015):	5.8
Nota d'Accés a la Universitat:	11.05

Assignatura a consultar:

Química II (240024)

Nota: usar un punt com a separador decimal

Aquesta és una aplicació de caràcter acadèmic.
Per a qualsevol dubte contacteu amb guillembalcells@hotmail.com

Figura 6.3. La interfície d'usuari en el cas d'una nota fora de rang.

APP pel càlcul estadístic de les notes

Premi Actualitzar per a veure la seva nota prevista de l'assignatura que seleccioni.

Àlgebra Lineal (240011):	6
Càlcul I (240012):	5.5
Mecànica Fonamental (240013):	5.2
Química I (240014):	6.3
Fonaments d'Informàtica (240015):	5.8
Nota d'Accés a la Universitat:	11.05

Assignatura a consultar:

Química II (240024)

La seva nota esperada de Química II és 5.96 i en un 95% de probabilitats que es trobarà dins del següent interval (2.97 ; 8.94)

Aquesta és una aplicació de caràcter acadèmic.
Per a qualsevol dubte contacteu amb guillembalcells@hotmail.com

Figura 6.4. La interfície d'usuari en el cas de totes les notes dins de rang.

A continuació es mostra el codi de la Funcio2. No obstant, abans convé recordar que en els models de regressió lineal múltiple, el valor previst de la variable resposta \hat{Y} per uns determinats valors de les variables explicatives X_z és:

$$\hat{Y} = X_z \hat{\beta} \quad (\text{Eq. 5.5})$$

I el corresponent interval de predicció de \hat{Y} és:

$$IP\ 100(1 - \alpha)\% = \hat{Y} \pm t_{(1-\alpha/2), (n-(p+1))} \hat{\sigma} \sqrt{1 + X_z^T (X^T X)^{-1} X_z} \quad (\text{Eq. 5.11})$$

S'observa que per els resultats acadèmics dels 1946 estudiants que s'estan usant, i les 6 notes que s'estan emprant com a variables explicatives del model, la dimensió de la matriu X és (1946 files x 6columnes). Conseqüentment, la inversa del producte $(X^T X)^{-1}$ és una matriu (6x6), els coeficients de la qual s'han calculat en R i s'han introduït al JavaScript a la matriu M1. Els valors a, b, c, d, e, f, són els corresponents components del vector $X_z^T (X^T X)^{-1}$.

La matriu M2 conté a cada fila, de la primera a la sisena columna, els valors dels coeficients respectius dels models de la regressió lineal múltiple, essent zero si la variable corresponent no forma part del model. La última columna conté la desviació estàndard estimada dels residus del model $\hat{\sigma}$, necessària pel càlcul de l'interval de predicció.

```
function Funcio2() {
    var a,b,c,d,e,f;

    M1 = [
        [1.043648e-04, -3.814867e-05, -4.413318e-05, -6.577035e-05, -4.804267e-05, -1.677814e-06],
        [-3.814867e-05, 3.769287e-04, -1.847927e-04, -6.837244e-05, -2.696540e-05, -1.501884e-05],
        [-4.413318e-05, -1.847927e-04, 4.301505e-04, -1.134869e-04, -2.948690e-05, -1.977127e-05],
        [-6.577035e-05, -6.837244e-05, -1.134869e-04, 5.124868e-04, -1.308659e-04, -5.060511e-05],
        [-4.804267e-05, -2.696540e-05, -2.948690e-05, -1.308659e-04, 3.002700e-04, -3.725650e-05],
        [-1.677814e-06, -1.501884e-05, -1.977127e-05, -5.060511e-05, -3.725650e-05, 1.250828e-04]
    ];

    a= notaA*M1[0][0] + nota1*M1[0][1] + nota2*M1[0][2] + nota3*M1[0][3] + nota4*M1[0][4] + nota5*M1[0][5];
    b= notaA*M1[1][0] + nota1*M1[1][1] + nota2*M1[1][2] + nota3*M1[1][3] + nota4*M1[1][4] + nota5*M1[1][5];
    c= notaA*M1[2][0] + nota1*M1[2][1] + nota2*M1[2][2] + nota3*M1[2][3] + nota4*M1[2][4] + nota5*M1[2][5];
    d= notaA*M1[3][0] + nota1*M1[3][1] + nota2*M1[3][2] + nota3*M1[3][3] + nota4*M1[3][4] + nota5*M1[3][5];
    e= notaA*M1[4][0] + nota1*M1[4][1] + nota2*M1[4][2] + nota3*M1[4][3] + nota4*M1[4][4] + nota5*M1[4][5];
    f= notaA*M1[5][0] + nota1*M1[5][1] + nota2*M1[5][2] + nota3*M1[5][3] + nota4*M1[5][4] + nota5*M1[5][5];

    M2 = [
        [0, 0.17365, 0.10130, 0.23829, 0.20059, 0.04501, 0.04331, 1.37],
        [0, 0.14480, 0.19654, 0.30916, 0.17936, 0, -0.04275, 1.543],
        [-2.56291, 0.34919, 0.36007, 0.14405, 0.11942, 0.04968, 0, 1.515],
        [0, 0.26822, 0.18557, 0.17113, 0, 0.12067, 0.03063, 1.521],
        [1.01971, 0.09657, 0.10178, 0.13681, 0.14271, 0, 0.24617, 1.775]
    ];
}
```

Figura 6.5. Primer part del codi de la Funcio2.

Seguidament, ja es pot calcular tant la nota esperada com el respectiu interval de predicció usant les equacions (Eq. 5.5) i (Eq 5.11) citades anteriorment. Si el límit inferior de l'interval de predicció és inferior a zero, passa a valer zero, i si el límit superior és superior a deu, passa a valer deu. En últim lloc, es mostren els resultats per pantalla.

```

if (assignatura == "Geometria") {
    notaCC = (M2[0][0] + M2[0][1]*notaA + M2[0][2]*nota1 + M2[0][3]*nota2 + M2[0][4]*nota3 + M2[0][5]*nota4 + M2[0][6]*nota5);
    CCmes = notaCC + 1.96*M2[0][7]*Math.sqrt(1 + (a*notaA + b*nota1 + c*nota2 + d*nota3 + e*nota4 + f*nota5));
    CCmenys = notaCC - 1.96*M2[0][7]*Math.sqrt(1 + (a*notaA + b*nota1 + c*nota2 + d*nota3 + e*nota4 + f*nota5));
} else if (assignatura == "Càlcul II") {
    notaCC = (M2[1][0] + M2[1][1]*notaA + M2[1][2]*nota1 + M2[1][3]*nota2 + M2[1][4]*nota3 + M2[1][5]*nota4 + M2[1][6]*nota5);
    CCmes = notaCC + 1.96*M2[1][7]*Math.sqrt(1 + (a*notaA + b*nota1 + c*nota2 + d*nota3 + f*nota5));
    CCmenys = notaCC - 1.96*M2[1][7]*Math.sqrt(1 + (a*notaA + b*nota1 + c*nota2 + d*nota3 + f*nota5));
} else if (assignatura == "Termodinàmica Fonamental") {
    notaCC = (M2[2][0] + M2[2][1]*notaA + M2[2][2]*nota1 + M2[2][3]*nota2 + M2[2][4]*nota3 + M2[2][5]*nota4 + M2[2][6]*nota5);
    CCmes = notaCC + 1.96*M2[2][7]*Math.sqrt(1 + (a*notaA + b*nota1 + c*nota2 + d*nota3 + e*nota4));
    CCmenys = notaCC - 1.96*M2[2][7]*Math.sqrt(1 + (a*notaA + b*nota1 + c*nota2 + d*nota3 + e*nota4));
} else if (assignatura == "Química II") {
    notaCC = (M2[3][0] + M2[3][1]*notaA + M2[3][2]*nota1 + M2[3][3]*nota2 + M2[3][4]*nota3 + M2[3][5]*nota4 + M2[3][6]*nota5);
    CCmes = notaCC + 1.96*M2[3][7]*Math.sqrt(1 + (a*notaA + b*nota1 + c*nota2 + e*nota4 + f*nota5));
    CCmenys = notaCC - 1.96*M2[3][7]*Math.sqrt(1 + (a*notaA + b*nota1 + c*nota2 + e*nota4 + f*nota5));
} else {
    notaCC = (M2[4][0] + M2[4][1]*notaA + M2[4][2]*nota1 + M2[4][3]*nota2 + M2[4][4]*nota3 + M2[4][5]*nota4 + M2[4][6]*nota5);
    CCmes = notaCC + 1.96*M2[4][7]*Math.sqrt(1 + (a*notaA + b*nota1 + c*nota2 + d*nota3 + f*nota5));
    CCmenys = notaCC - 1.96*M2[4][7]*Math.sqrt(1 + (a*notaA + b*nota1 + c*nota2 + d*nota3 + f*nota5));
}

CCmenys = Math.max(0, CCmenys.toFixed(2));
CCmes = Math.min(10, CCmes.toFixed(2));
notaCC = notaCC.toFixed(2);

document.getElementById("resultat").innerHTML = "La seva nota esperada de " + assignatura + " és " + notaCC +
" i en un 95% de probabilitats que es trobarà dins del següent interval (" + CCmenys + " ; " + CCmes + ") " + "<br>" + "<br>" + "<br>";
}

```

Figura 6.6. Segona part del codi de la Funcio2.

Finalment, es va demanar a una amiga, de nom Sandra Orozco, estudiant de la Facultat de Matemàtiques i Estadística de la UPC, de poder penjar la pàgina web al servidor que se'ls ofereix als estudiants de dita facultat. L'enllaç per a accedir a la pàgina web és el següent: <http://fourier.upc.es/~sandra.orozco/guilem.balcellsTFG.html>

7. PROPOSTA DE CONTINUACIÓ DEL TREBALL

Aquest treball ofereix la possibilitat de ser continuat estenent les funcionalitats de l'aplicació web. Les propostes de continuació recomanades són les següents:

- a) Ampliar el rang d'assignatures a preveure a tot el grau.
- b) Ampliar el rang d'assignatures amb les que realitza la regressió lineal múltiple a totes les assignatures cursades anteriorment per l'estudiant usuari de l'aplicació.
- c) Ampliar el model de regressió lineal múltiple afegint variables explicatives que no siguin resultats acadèmics, com per exemple el quadrimestre en que es cursa (primavera o tardor), el coordinador, si l'estudiant viu o no a Barcelona, si l'estudiant prové de batxillerat públic o privat, etc.
- d) Permetre a l'usuari la possibilitat de triar el % de l'interval de confiança.
- e) Automatitzar el procés de creació de les matrius M1 i M2 del Java Script de manera que els coeficients no s'hagin d'escriure a ma, sinó que s'importin automàticament de R.

8. PLANIFICACIÓ TEMPORAL I COSTOS

8.1. Planificació temporal

En primer lloc, es van revisar els conceptes adquirits a l'assignatura de Fonaments d'Informàtica, d'Informàtica i Estadística. Seguidament, es va procedir a escollir amb quins codis i programes es treballaria durant al llarg del treball (R, HTML, CSS, JavaScript), al mateix temps que s'ampliaven els coneixements d'estadística. A continuació, es va procedir a aprendre els codis de programació web. Al mateix temps es va començar a redactar la memòria.

Un cop adquirits els coneixements necessaris, es va procedir a elaborar els algorismes amb R: el filtrat de les dades, els models inicials, l'eliminació progressiva i l'anàlisi dels models. També es va començar a elaborar l'aplicació web.

Seguidament, es va acabar el codi en R i després es va completar el codi web. Finalment, es va també finalitzar la memòria.

A continuació es presenta el diagrama de Gantt seguit per a l'elaboració del projecte.

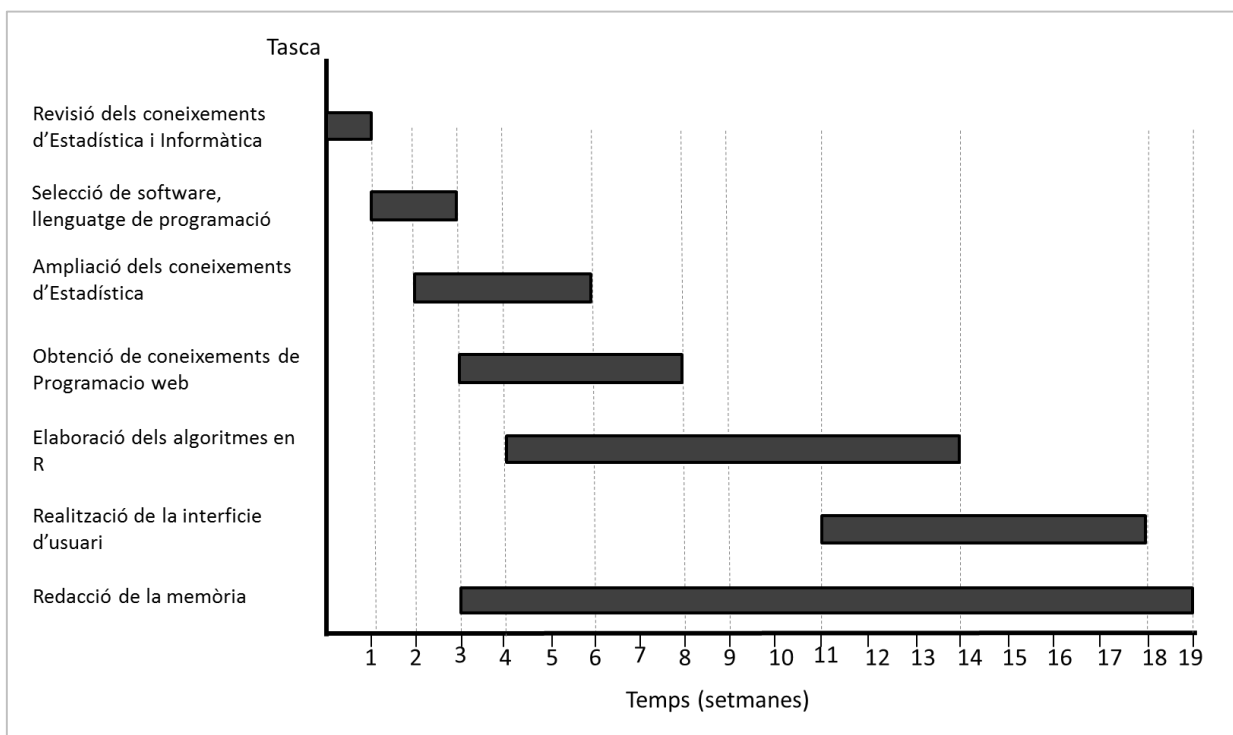


Figura 8.1. Diagrama de Gantt del projecte.

8.2. Costos

L'abast del treball i el marcat caràcter teòric d'aquest dificulten una precisa comptabilització de la despesa d'aquet. Així doncs, es procedirà tant sols a comptabilitzar els costos principals:

En primer lloc es tindrà en compte el treball realitzat per l'estudiant en qüestió, assimilant el seu sou al d'un graduat en enginyeria de perfil junior. El sou estàndard de mercat d'un enginyer graduat junior és d'uns 20€/hora aproximadament. D'altra banda es considera que el conjunt d'hores treballades és l'equivalent a 12 crèdits ECTS –cada crèdit són de mitjana 28 hores de treball- i per tant 336 hores. Així doncs, el cost de l'estudiant és de 336 hores a raó de 20€/hora: 6.720€.

En segon lloc, la feina de supervisió realitzada pel tutor del projecte, assimilant el seu sou al d'un enginyer sènior. S'aproxima el temps dedicat a 30 hores i el sou de l'enginyer a 40€/l'hora, i per tant una despesa total de 1.200€.

El cost en llicències dels programes de Microsoft Office emprats. D'altra banda, R és gratuït. Així doncs, el cost de quatre anys de llicències de Microsoft Office per a universitaris és de 79,90€, i per tant la despesa corresponent a 4 mesos d'ús és de:

$$\text{cost llicències} = \frac{79,90\text{€}}{4 \text{ anys}} \times \frac{1 \text{ any}}{12 \text{ mesos}} \times 4 \text{ mesos} = 6,66\text{€}$$

El cost de l'ordinador es comptabilitza considerant una esperança de vida de 5 anys -60 mesos- i un preu total de 600€:

$$\text{cost ordinador} = \frac{600\text{€}}{60 \text{ mesos}} \times 4 \text{ mesos} = 40\text{€}$$

El consum de llum, internet i telèfon emprats és aproximadament de 400€.

El cost en transport i material d'oficina es comptabilitza en 100€.

Així doncs, el cost total del projecte és la suma de les xifres anteriors:

$$\text{cost total} = 6.720 + 1.200 + 6,66 + 40 + 400 + 100 = 8.466,66\text{€}$$

A continuació es mostra una taula resum de les despeses:

Concepte del cost	Cost (€)
Hores estudiant	6.720
Hores tutor	1.200
Llicències	6,66
Ordinador	40
Llum, internet i telèfon	400
Transport i material d'oficina	100
Cost Total	8.466,66

Taula 8.1. Taula resum de les despeses.

9. IMPACTE SOCIAL I AMBIENTAL

9.1. Impacte social

L'aplicació web desenvolupada en aquest treball permetrà als estudiants de primer curs del Grau en Enginyeria en Tecnologies Industrials la possibilitat de conèixer, al acabar el primer quadrimestre de grau, quines son les assignatures del segon quadrimestre que els resultaran més i menys senzilles d'encarar. D'aquesta manera, l'estudiant podrà organitzar-se millor acadèmicament.

A més, la possible ampliació de l'aplicació web, seguint les recomanacions del punt 7 d'aquest informe (Proposta de Continuació del Treball), afinaria la qualitat dels resultats i permetria als estudiants de poder emprar la pàgina web al llarg de tot el grau.

9.2. Impacte ambiental

Degut a la naturalesa d'aquest projecte, el qual és de caràcter informàtic, l'impacte ambiental que presenta és molt baix. Les principals causes d'impacte ambiental són:

- a) el consum energètic d'electricitat que necessita l'ordinador per tal d'estar operatiu, ja que aquest ha estat llargs períodes de temps en funcionament.
- b) La utilització dels elements de climatització de la sala on s'ha treballat.
- c) L'efecte mediambiental dels residus generats (papers, bolígrafs, etc.)

CONCLUSIONS

En aquest projecte s'ha desenvolupat una aplicació web que permet a l'usuari conèixer les previsions de les notes del segon quadrimestre del Grau en Enginyeria en Tecnologies Industrials en funció de les notes del primer quadrimestre.

D'acord amb els objectius del projecte, s'han filtrat les dades per a obtenir aquelles que eren d'interès en un format òptim per a treballar. S'ha emprat la regressió lineal múltiple com a mètode de predicció, i l'eliminació progressiva com a mètode per a optimitzar el model. S'han verificat les hipòtesis dels models i també s'han analitzat per extreure'n informació. Seguidament s'ha desenvolupat l'aplicació web mitjançant HTML, CSS i JavaScript.

S'observa que l'aplicació ha assolit els objectius. No obstant això, el model de regressió no reflexa la variabilitat de les notes en la seva totalitat i aquest fet es reflexa en que l'interval de confiança de les previsions abasta un rang considerablement gran.

Finament, s'han fet algunes propostes d'ampliació i millora de l'aplicació.

AGRAÏMENTS

M'agradaria donar les gràcies al meu tutor, en Lluís Solano, per l'ajuda que he rebut en cadascuna de les etapes d'aquest treball. Agraeixo també tot el suport de la meua família i dels meus amics, i especialment a la Sandra Orozco, qui m'ha ajudat a afrontar els reptes que se'm han anat plantejant al llarg del desenvolupament del projecte.

BIBLIOGRAFIA

Referències bibliogràfiques

1. **The R Foundation.** *The R Project for Statistical Computing*. [En línia] [Data: 17 / Abril / 2016.] <https://www.r-project.org/>.
2. **Refsnes Data.** *w3school.com The World's Largest Web Developer Site*. [En línia] [Data: 17 / Abril / 2016.] <http://www.w3schools.com/default.asp>.
3. **Madsen, Henrik.** *Time Series Analysis*. Copenhagen : Chapman & Hall, 2008.
4. **Andersen, Elisabeth W.** *Statistical modelling: Theory and practice, Week 2. Multiple Regression*. Copenhagen : DTU Applied Mathematics and Computer Science, September / 2015.

Bibliografia complementària

1. **Peter Dalgaard.** *Introductory Statistics with R*. Copenhagen: Springer, 2008.
2. **Simon J. Sheather.** *A Modern Approach to Regression with R*. Texas: Springer, 2009.
3. **N.H. Bingham, John M. Fry.** *Regression. Linear Models in Statistics*. London: Springer, 2010.